

Texts in Applied Mathematics 56

Mark H. Holmes

# Introduction to the Foundations of Applied Mathematics

*Second Edition*



Springer

# Texts in Applied Mathematics

Volume 56

## Editors-in-chief

A. Bloch, University of Michigan, Ann Arbor, USA  
C. L. Epstein, University of Pennsylvania, Philadelphia, USA  
A. Goriely, University of Oxford, Oxford, UK  
L. Greengard, New York University, New York, USA

## Series Editors

J. Bell, Lawrence Berkeley National Lab, Berkeley, USA  
R. Kohn, New York University, New York, USA  
P. Newton, University of Southern California, Los Angeles, USA  
C. Peskin, New York University, New York, USA  
R. Pego, Carnegie Mellon University, Pittsburgh, USA  
L. Ryzhik, Stanford University, Stanford, USA  
A. Singer, Princeton University, Princeton, USA  
A. Stevens, Universität Münster, Münster, Germany  
A. Stuart, University of Warwick, Coventry, UK  
T. Witelski, Duke University, Durham, USA  
S. Wright, University of Wisconsin, Madison, USA

The mathematization of all sciences, the fading of traditional scientific boundaries, the impact of computer technology, the growing importance of computer modeling and the necessity of scientific planning all create the need both in education and research for books that are introductory to and abreast of these developments. The aim of this series is to provide such textbooks in applied mathematics for the student scientist. Books should be well illustrated and have clear exposition and sound pedagogy. Large number of examples and exercises at varying levels are recommended. TAM publishes textbooks suitable for advanced undergraduate and beginning graduate courses, and complements the Applied Mathematical Sciences (AMS) series, which focuses on advanced textbooks and research-level monographs.

More information about this series at <http://www.springer.com/series/1214>

Mark H. Holmes

# Introduction to the Foundations of Applied Mathematics

Second Edition

 Springer



Mark H. Holmes  
Department of Mathematical Sciences  
Rensselaer Polytechnic Institute  
Troy, NY, USA

ISSN 0939-2475 ISSN 2196-9949 (electronic)  
Texts in Applied Mathematics  
ISBN 978-3-030-24260-2 ISBN 978-3-030-24261-9 (eBook)  
<https://doi.org/10.1007/978-3-030-24261-9>

Mathematics Subject Classification (2010): Primary: 76Axx, 76Bxx, 76Dxx, 74Bxx, 74Dxx, 74Hxx, 74Jxx, 74Axx, 34D05, 34E05, 34E10, 34E13, 35C06, 35C07, 35F50, 60J60, 60J65

1<sup>st</sup> edition: © Springer Science+Business Media, LLC 2009  
© Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

*To Colette, Matthew, and Marianna*

## Preface to the Second Edition

The principal changes are directed to improving the presentation of the material. This includes rewriting and reorganizing certain sections, adding new examples, and reorganizing and embellishing the exercises. The added examples range from the relatively minor to the more extensive, such as the added material for water waves. This edition also provided an opportunity to update the references.

Another reason for this edition concerns the changes in publishing over the last decade. The improvements in digital books and the interest in students for having an ebook version were motivating reasons for working on a new edition.

Finally, the one or two typos in the first edition were also corrected, and thanks go to Ash Kapila, Emily Fagerstrom, Jan Medlock, and Kevin DelBene for finding them.

Troy, NY, USA  
March 2019

Mark H. Holmes

# Preface to the First Edition

FOAM. This acronym has been used for over 50 years at Rensselaer to designate an upper-division course entitled, Foundations of Applied Mathematics. This course was started by George Handelman in 1956, when he came to Rensselaer from the Carnegie Institute of Technology. His objective was to closely integrate mathematical and physical reasoning, and in the process enable students to obtain a qualitative understanding of the world we live in. FOAM was soon taken over by a young faculty member, Lee Segel. About this time a similar course, Introduction to Applied Mathematics, was introduced by Chia-Ch'iao Lin at the Massachusetts Institute of Technology. Together Lin and Segel, with help from Handelman, produced one of the landmark textbooks in applied mathematics, *Mathematics Applied to Deterministic Problems in the Natural Sciences*. This was originally published in 1974, and republished in 1988 by the Society for Industrial and Applied Mathematics, in their Classics Series.

This textbook comes from the author teaching FOAM over the last few years. In this sense, it is an updated version of the Lin and Segel textbook. The objective is definitely the same, which is the construction, analysis, and interpretation of mathematical models to help us understand the world we live in. However, there are some significant differences. Lin and Segel, like many recent modeling books, is based on a case study format. This means that the mathematical ideas are introduced in the context of a particular application. There are certainly good reasons why this is done, and one is the immediate relevance of the mathematics. There are also disadvantages, and one pointed out by Lin and Segel is the fragmentary nature of the development. However, there is another, more important reason for not following a case studies approach. Science evolves, and this means that the problems of current interest continually change. What does not change as quickly is the approach used to derive the relevant mathematical models, and the methods used to analyze the models. Consequently, this book is written in such a way as to establish the mathematical ideas underlying model development independently of a specific application. This does not mean applications are not considered, they are, and connections with experiment are a staple of this book.

The first two chapters establish some of the basic mathematical tools that are needed. The model development starts in Chap. 3, with the study of kinetics. The goal of this chapter is to understand how to model interacting populations. This does not account for the spatial motion of the populations, and this is the objective of Chaps. 4 and 5. What remains is to account for the forces in the system, and this is done in Chap. 6. The last three chapters concern the application to specific problems and the generalization of the material to more geometrically realistic systems. The book, as well as the individual chapters, is written in such a way that the material becomes more sophisticated as you progress. This provides some flexibility in how the book is used, allowing consideration for the breadth and depth of the material covered.

The principal objective of this book is the derivation and analysis of mathematical models. Consequently, after deriving a model, it is necessary to have a way to solve the resulting mathematical problem. A few of the methods developed here are standard topics in upper-division applied math courses, and in this sense there is some overlap with the material covered in those courses. Examples are the Fourier and Laplace transforms, and the method of characteristics. On the other hand, other methods that are used here are not standard, and this includes perturbation approximations and similarity solutions. There are also unique methods, not found in traditional textbooks, that rely on both the mathematical and physical characteristics of the problem.

The prerequisite for this text is a lower-division course in differential equations. The implication is that you have also taken two or three semesters of calculus, which includes some component of matrix algebra. The one topic from calculus that is absolutely essential is Taylor's theorem, and for this reason a short summary is included in the appendix. Some of the more sophisticated results from calculus, related to multidimensional integral theorems, are not needed until Chap. 8.

To learn mathematics you must work out problems, and for this reason the exercises in the text are important. They vary in their difficulty, and cover most of the topics in the chapter. Some of the answers are available, and can be found at [www.holmes.rpi.edu](http://www.holmes.rpi.edu). This web page also contains a typos list.

I would like to express my gratitude to the many students who have taken my FOAM course at Rensselaer. They helped me immeasurably in understanding the subject, and provided much-needed encouragement to write this book. It is also a pleasure to acknowledge the suggestions of John Ringland, and his students, who read an early version of the manuscript.

Troy, NY, USA  
March 2009

Mark H. Holmes

# Contents

<b>1</b>	<b>Dimensional Analysis</b>	1
1.1	Introduction	1
1.2	Examples of Dimensional Reduction	3
1.2.1	Maximum Height of a Projectile	5
1.2.2	Drag on a Sphere	6
1.2.3	Toppling Dominoes	14
1.2.4	Endnotes	16
1.3	Theoretical Foundation	16
1.3.1	Pattern Formation	20
1.4	Similarity Variables	22
1.4.1	Dimensional Reduction	23
1.4.2	Similarity Solution	25
1.5	Nondimensionalization and Scaling	27
1.5.1	Projectile Problem	27
1.5.2	Weakly Nonlinear Diffusion	31
1.5.3	Endnotes	34
	Exercises	34
<b>2</b>	<b>Perturbation Methods</b>	49
2.1	Regular Expansions	49
2.2	How to Find a Regular Expansion	53
2.2.1	Given a Specific Function	53
2.2.2	Given an Algebraic or Transcendental Equation	56
2.2.3	Given an Initial Value Problem	60
2.3	Scales and Approximation	64
2.4	Introduction to Singular Perturbations	66
2.5	Introduction to Boundary Layers	69
2.5.1	Endnotes	76
2.6	Examples Involving Boundary Layers	77
2.6.1	Example 1: Layer at Left End	77
2.6.2	Example 2: Layer at Right End	79

2.6.3	Example 3: Boundary Layer at Both Ends	81
2.7	Multiple Scales	84
2.7.1	Regular Expansion	85
2.7.2	Multiple Scales Expansion	88
	Exercises	92
<b>3</b>	<b>Kinetics</b>	103
3.1	Introduction	103
3.1.1	Radioactive Decay	103
3.1.2	Predator-Prey	104
3.1.3	Epidemic Model	104
3.2	Kinetic Equations	105
3.2.1	The Law of Mass Action	107
3.2.2	Conservation Laws	109
3.2.3	Steady States	111
3.2.4	Examples	111
3.2.5	End Notes	113
3.3	Modeling Using the Law of Mass Action	114
3.3.1	Michaelis-Menten Kinetics	115
3.3.2	Disease Modeling	116
3.3.3	Reverse Mass Action	118
3.4	General Mathematical Formulation	119
3.5	Steady States and Stability	123
3.5.1	Reaction Analysis	123
3.5.2	Geometric Analysis	124
3.5.3	Perturbation Analysis	126
3.6	Solving the Michaelis-Menten Problem	134
3.6.1	Numerical Solution	134
3.6.2	Quasi-Steady-State Approximation	135
3.6.3	Perturbation Approach	137
3.7	Oscillators	143
3.7.1	Stability	145
3.8	Modeling with the QSSA	148
3.9	Epilogue	151
	Exercises	151
<b>4</b>	<b>Diffusion</b>	165
4.1	Introduction	165
4.2	Random Walks and Brownian Motion	167
4.2.1	Calculating $w(m, n)$	170
4.2.2	Large $n$ Approximation	172
4.3	Continuous Limit	174
4.3.1	What Does $D$ Signify?	175
4.4	Solutions of the Diffusion Equation	178
4.4.1	Point Source Solution	178
4.4.2	A Step Function Initial Condition	183

4.5	Fourier Transform .....	186
4.5.1	Transformation of Derivatives .....	187
4.5.2	Convolution Theorem .....	189
4.5.3	Solving the Diffusion Equation .....	191
4.6	Continuum Formulation of Diffusion .....	194
4.6.1	Balance Law .....	195
4.6.2	Fick's Law of Diffusion .....	196
4.6.3	Reaction-Diffusion Equations .....	203
4.7	Random Walks and Diffusion in Higher Dimensions .....	205
4.7.1	Diffusion Equation .....	207
4.8	Langevin Equation .....	211
4.8.1	Properties of the Random Forcing .....	213
4.8.2	Endnotes .....	219
	Exercises .....	220
<b>5</b>	<b>Traffic Flow .....</b>	<b>233</b>
5.1	Introduction .....	233
5.2	Continuum Variables .....	233
5.2.1	Density .....	234
5.2.2	Flux .....	236
5.3	Balance Law .....	237
5.3.1	Velocity Formulation .....	238
5.4	Constitutive Laws .....	239
5.4.1	Constant Velocity .....	241
5.4.2	Linear Velocity: Greenshields Law .....	241
5.4.3	General Velocity Formulation .....	242
5.4.4	Flux and Velocity .....	244
5.4.5	Reality Check .....	244
5.5	Constant Velocity .....	245
5.5.1	Characteristics .....	248
5.6	Density Dependent Velocity .....	252
5.6.1	Small Disturbance Approximation .....	253
5.6.2	Method of Characteristics .....	255
5.6.3	Rankine-Hugoniot Condition .....	260
5.6.4	Shock Waves .....	262
5.6.5	Expansion Fan .....	264
5.6.6	Summary .....	270
5.6.7	Additional Examples .....	271
5.7	Cellular Automata Modeling .....	276
	Exercises .....	282
<b>6</b>	<b>Continuum Mechanics: One Spatial Dimension .....</b>	<b>295</b>
6.1	Introduction .....	295
6.2	Frame of Reference .....	295
6.2.1	Material Coordinates .....	296
6.2.2	Spatial Coordinates .....	297



6.2.3	Material Derivative	300
6.2.4	End Notes	302
6.3	Mathematical Tools	304
6.4	Continuity Equation	305
6.4.1	Material Coordinates	306
6.5	Momentum Equation	307
6.5.1	Material Coordinates	309
6.6	Summary of the Equations of Motion	309
6.7	Steady-State Solution	311
6.8	Constitutive Law for an Elastic Material	312
6.8.1	Derivation of Strain	314
6.8.2	Material Linearity	316
6.8.3	Material Nonlinearity	319
6.8.4	End Notes	320
6.9	Morphological Basis for Deformation	321
6.9.1	Metals	321
6.9.2	Elastomers	324
6.10	Restrictions on Constitutive Laws	325
6.10.1	Frame-Indifference	326
6.10.2	Entropy Inequality	328
6.10.3	Hyperelasticity	332
	Exercises	335
<b>7</b>	<b>Elastic and Viscoelastic Materials</b>	<b>345</b>
7.1	Linear Elasticity	345
7.1.1	Method of Characteristics	348
7.1.2	Laplace Transform	350
7.1.3	Geometric Linearity	362
7.2	Viscoelasticity	363
7.2.1	Mass, Spring, Dashpot Systems	363
7.2.2	Equations of Motion	366
7.2.3	Integral Formulation	370
7.2.4	Generalized Relaxation Functions	372
7.2.5	Solving Viscoelastic Problems	373
	Exercises	377
<b>8</b>	<b>Continuum Mechanics: Three Spatial Dimensions</b>	<b>389</b>
8.1	Introduction	389
8.2	Material and Spatial Coordinates	390
8.2.1	Deformation Gradient	392
8.3	Material Derivative	395
8.4	Mathematical Tools	397
8.4.1	General Balance Law	400
8.4.2	Direct Notation and Tensors	401
8.5	Continuity Equation	401
8.5.1	Incompressibility	402

8.6	Linear Momentum Equation .....	403
8.6.1	Stress Tensor .....	404
8.6.2	Differential Form of Equation .....	406
8.7	Angular Momentum .....	407
8.8	Summary of the Equations of Motion .....	407
8.8.1	The Assumption of Incompressibility .....	408
8.9	Constitutive Laws .....	409
8.9.1	Representation Theorem and Invariants .....	413
8.10	Newtonian Fluid .....	414
8.10.1	Pressure .....	415
8.10.2	Viscous Stress .....	415
8.11	Equations of Motion for a Viscous Fluid .....	418
8.11.1	Incompressibility .....	419
8.11.2	Boundary and Initial Conditions .....	420
8.12	Material Equations of Motion .....	423
8.12.1	Frame-Indifference .....	426
8.12.2	Elastic Solid .....	426
8.12.3	Linear Elasticity .....	429
8.13	Energy Equation .....	430
8.13.1	Incompressible Viscous Fluid .....	431
8.13.2	Elasticity .....	432
	Exercises .....	434
<b>9</b>	<b>Newtonian Fluids</b> .....	<b>445</b>
9.1	Steady Flow .....	446
9.1.1	Plane Couette Flow .....	446
9.1.2	Poiseuille Flow .....	450
9.2	Vorticity .....	453
9.2.1	Vortex Motion .....	455
9.3	Irrotational Flow .....	456
9.3.1	Potential Flow .....	459
9.4	Ideal Fluid .....	461
9.4.1	Circulation and Vorticity .....	462
9.4.2	Potential Flow .....	465
9.4.3	End Notes .....	469
9.5	Boundary Layers .....	469
9.5.1	Impulsive Plate .....	470
9.5.2	Blasius Boundary Layer .....	471
9.6	Water Waves .....	477
9.6.1	Interface Conditions .....	478
9.6.2	Traveling Waves .....	479
9.6.3	Wave Generation .....	481
	Exercises .....	486

<b>A Taylor's Theorem</b> .....	497
A.1 Single Variable .....	497
A.1.1 Simplification via Substitution .....	498
A.2 Two Variables .....	499
A.3 Multivariable Versions .....	500
<b>B Fourier Analysis</b> .....	503
B.1 Fourier Series .....	503
B.2 Fourier Transform .....	505
<b>C Stochastic Differential Equations</b> .....	507
<b>D Identities</b> .....	509
D.1 Trace .....	509
D.2 Determinant .....	509
D.3 Vector Calculus .....	510
D.4 Miscellaneous .....	510
<b>E Equations for a Newtonian Fluid</b> .....	511
E.1 Cartesian Coordinates .....	511
E.2 Cylindrical Coordinates .....	511
<b>References</b> .....	515
<b>Index</b> .....	523

# Chapter 1

## Dimensional Analysis



### 1.1 Introduction

Before beginning the material on dimensional analysis, it is worth considering a simple example that demonstrates what we are doing. One that qualifies as simple is the situation of when an object is thrown upwards. The resulting mathematical model for this is an equation for the height  $x(t)$  of the projectile from the surface of the Earth at time  $t$ . This equation is determined using Newton's second law,  $F = ma$ , and the law of gravitation. The result is

$$\frac{d^2x}{dt^2} = -\frac{gR^2}{(R+x)^2}, \quad \text{for } 0 < t, \quad (1.1)$$

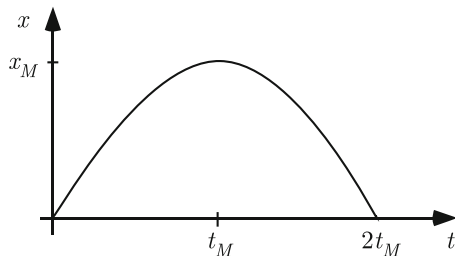
where  $g$  is the gravitational acceleration constant and  $R$  is the radius of the Earth. Finding the solution  $x$  of this equation requires two integrations. Each will produce an integration constant, and we need more information to find these constants. This is done by specifying the initial conditions. Assuming the projectile starts at the surface with velocity  $v_0$ , then the initial conditions are as follows:

$$x(0) = 0, \quad (1.2)$$

$$\frac{dx}{dt}(0) = v_0. \quad (1.3)$$

The resulting initial value problem for  $x$  consists in finding the solution of (1.1) that satisfies (1.2) and (1.3). Mathematically, the problem is challenging because it involves solving a second-order nonlinear differential equation. One option for finding the solution is simply to use a computer. However, the limitation with this is that it does not provide much insight into how the solution depends on the terms in the equation. One of the primary objectives of this text is to use mathematics to derive a fundamental understanding of how and why things work the way they do,

**Fig. 1.1** The solution (1.5) of the projectile problem in a uniform gravitational field



and so, we are very interested in obtaining at least an approximate solution of this problem. This is the same point-of-view taken in most physics books and it is worth looking at how they might address this issue.

Adopting, for the moment, the typical Physics I approach, in looking at the equation in (1.1) it is not unreasonable to assume  $R$  is significantly larger than even the largest value of  $x$ . If true, then we should be able to replace the  $x + R$  term with just  $R$ . In this case, the problem reduces to solving

$$\frac{d^2x}{dt^2} = -g, \quad \text{for } 0 < t. \quad (1.4)$$

Integrating and then using the two initial conditions yields

$$x(t) = -\frac{1}{2}gt^2 + v_0t. \quad (1.5)$$

This solution is shown schematically in Fig. 1.1. We have what we wanted, a relatively simple expression that serves as an approximation to the original nonlinear problem. To complete the derivation we should check that the assumption made in the derivation is satisfied, namely  $x$  is much smaller than  $R$ . Now, the maximum height for (1.5) occurs when

$$\frac{dx}{dt} = 0. \quad (1.6)$$

Solving this equation yields  $t = v_0/g$  and from this it follows that the maximum height is

$$x_M = \frac{v_0^2}{2g}. \quad (1.7)$$

Therefore, we must require that  $v^2/(2g)$  is much less than  $R$ , which we write as  $v_0^2/(2g) \ll R$ .

It is now time to critique the above derivation. The first criticism is that the approach is heuristic. The reason is that even though the argument for replacing  $x + R$  with  $R$  seems plausible, we simply ignored a particular term in the equation.

The projectile problem is not particularly complicated, so dropping a term as we did is straightforward. However, in the real world where problems can be quite complicated, dropping a term in one part of the problem can lead to inconsistencies in another part. A second criticism can be made by asking a question. Specifically, what exactly is the effect of the nonlinearity on the projectile? Our reduction replaced the nonlinear gravitational force, which is the right-hand side of (1.1), with a uniform gravitational field given by  $-g$ . Presumably if gravity decreases with height, then the projectile will be going higher than we would expect based on our approximation in (1.5). It is of interest to understand quantitatively what this nonlinear effect is and whether it might interfere with our reduction.

Based on the comments of the previous paragraph we need to make the reduction process more systematic. The procedure that is used to simplify the problem should enable us to know exactly what is large or small in the problem, and it should also enable us to construct increasingly more accurate approximations to the problem. Explaining what is involved in a systematic reduction occurs in two steps. The first, which is the objective of this chapter, involves the study of dimensions and how these can be used to simplify the mathematical formulation of the problem. After this, in Chap. 2, we develop techniques to construct accurate approximations of the resulting equations.

## 1.2 Examples of Dimensional Reduction

The first idea that we explore will, on the surface, seem to be rather simple, but it is actually quite profound. It has to do with the dimensions of the physical variables, or parameters, in a problem. To illustrate, suppose we know that the speed  $s$  of a ball is determined by its radius  $r$  and the length of time  $t$  it has been moving. Implicit in this statement is the assumption that the speed does not depend on any other physical variable. In mathematical terms we have that  $s = f(r, t)$ . The function  $f$  is not specified and all we know is that there is some expression that connects the speed with  $r$  and  $t$ . The only possible way to combine these two quantities to produce the dimension of speed is through their ratio  $r/t$ . For example, it is impossible to have  $s = \alpha r + \beta t$  without  $\alpha$  and  $\beta$  having dimensions. This would mean  $\alpha$  and  $\beta$  are physical parameters, and we have assumed there are no others in the problem. This observation enables us to conclude that based on the original assumptions that the only function we can have is  $s = \alpha r/t$ , where  $\alpha$  is a number.

What we are seeing in this example is that the dimensions of the variables in the problem end up dictating the form of the function. This is a very useful information and we will spend some time exploring how to exploit this idea. To set the stage, we need to introduce the needed terminology and notation.

First, there is the concept of a fundamental dimension. As is well known, physical variables such as force, density, and velocity can be broken down into length  $L$ , time  $T$ , and mass  $M$  (see Table 1.1). Moreover, length, time, and mass are independent in the sense that one of them cannot be written in terms of the other two. For these two

**Table 1.1** Fundamental dimensions for commonly occurring quantities. A quantity with a one in the dimensions column is dimensionless

Quantity	Dimensions	Quantity	Dimensions
Acceleration	$LT^{-2}$	Enthalpy	$ML^2T^{-2}$
Angle	1	Entropy	$ML^2T^{-2}\theta^{-1}$
Angular acceleration	$T^{-2}$	Gas constant	$ML^2T^{-2}\theta^{-1}$
Angular momentum	$ML^2T^{-1}$	Internal energy	$ML^2T^{-2}$
Angular velocity	$T^{-1}$	Specific heat	$L^2T^{-2}\theta^{-1}$
Area	$L^2$	Temperature	$\theta$
Energy, work	$ML^2T^{-2}$	Thermal conductivity	$MLT^{-3}\theta^{-1}$
Force	$MLT^{-2}$	Thermal diffusivity	$L^2T^{-1}$
Frequency	$T^{-1}$	Heat transfer coefficient	$MT^{-3}\theta^{-1}$
Concentration	$L^{-3}$		
Length	$L$	Capacitance	$M^{-1}L^{-2}T^4I^2$
Mass	$M$	Charge	TI
Mass density	$ML^{-3}$	Charge density	$L^{-3}TI$
Momentum	$MLT^{-1}$	Electrical conductivity	$M^{-1}L^{-3}T^3I^2$
Power	$ML^2T^{-3}$	Admittance	$L^{-2}M^{-1}T^3I^2$
Pressure, stress, elastic modulus	$ML^{-1}T^{-2}$	Electric potential, voltage	$ML^2T^{-3}I^{-1}$
Surface tension	$MT^{-2}$	Current density	$L^{-2}I$
Time	$T$	Electric current	$I$
Torque	$ML^2T^{-2}$	Electric field intensity	$MLT^{-3}I^{-1}$
Velocity	$LT^{-1}$	Inductance	$ML^2T^{-2}I^{-2}$
Viscosity (dynamic)	$ML^{-1}T^{-1}$	Magnetic intensity	$L^{-1}I$
Viscosity (kinematic)	$L^2T^{-1}$	Magnetic flux density	$MT^{-2}I^{-1}$
Volume	$L^3$	Magnetic permeability	$MLT^{-2}I^{-2}$
Wave length	$L$	Electric permittivity	$M^{-1}L^{-3}T^4I^2$
Strain	1	Electric resistance	$ML^2T^{-3}I^{-2}$

reasons we will consider  $L$ ,  $T$ , and  $M$  as *fundamental dimensions*. For problems involving thermodynamics we will expand this list to include temperature ( $\theta$ ) and for electrical problems we add current ( $I$ ). This gives rise to the following.

**Dimensions Notation.** *Given a physical quantity  $q$ , the fundamental dimensions of  $q$  will be denoted as  $\llbracket q \rrbracket$ . In the case of when  $q$  is dimensionless,  $\llbracket q \rrbracket = 1$ .*

So, for example, from the projectile problem,  $\llbracket v_0 \rrbracket = L/T$ ,  $\llbracket x \rrbracket = L$ ,  $\llbracket g \rrbracket = L/T^2$ , and  $\llbracket x_M/R \rrbracket = 1$ .

It is important to understand that nothing is being assumed about which specific system of units is used to determine the values of the variables or parameters. Dimensional analysis requires that the equations be independent of the system of units. For example, both Newton's law  $F = ma$  and the differential equation (1.1) do not depend on the specific system one selects. For this reason these equations are said to be *dimensionally homogeneous*. If one were to specialize (1.1) to SI units

and set  $R = 6378 \text{ km}$  and  $g = 9.8 \text{ m/s}^2$  they would end up with an equation that is not dimensionally homogeneous.

### 1.2.1 Maximum Height of a Projectile

The process of dimensional reduction will be explained by applying it to the projectile problem. To set the stage, suppose we are interested in the maximum height  $x_M$  of the projectile as shown in Fig. 1.1. For a uniform gravitational field the force is  $F = -mg$ . With this, and given the initial conditions in (1.2) and (1.3), it is assumed that the only physical parameters that  $x_M$  depends on are  $g$ ,  $v_0$ , and the mass  $m$  of the projectile. Mathematically this assumption is written as  $x_M = f(g, m, v_0)$ . The function  $f$  is unknown but we are going to see if the dimensions can be used to simplify the expression.

The only way to combine  $g$ ,  $m$ , and  $v_0$  to produce the correct dimensions is through a product or ratio. So, our start-off hypothesis is that there are numbers  $a$ ,  $b$ , and  $c$ , so that

$$[x_M] = [m^a v_0^b g^c]. \quad (1.8)$$

Using the fundamental dimensions for these variables the above equation is equivalent to

$$\begin{aligned} L &= M^a (L/T)^b (L/T^2)^c \\ &= M^a L^{b+c} T^{-b-2c}. \end{aligned} \quad (1.9)$$

Equating the exponents of the respective terms in this equation we conclude

$$\begin{aligned} L : \quad & b + c = 1, \\ T : \quad & -b - 2c = 0, \\ M : \quad & a = 0. \end{aligned}$$

Solving these equations we obtain  $a = 0$ ,  $b = 2$ , and  $c = -1$ . This means the only way to produce the dimensions of length using  $m$ ,  $v_0$ , and  $g$  is through the ratio  $v_0^2/g$ . Given our start-off assumption (1.8), we conclude that  $x_M$  is proportional to  $v_0^2/g$ . In other words, the original assumption that  $x_M = f(g, m, v_0)$  dimensionally reduces to the expression

$$x_M = \alpha \frac{v_0^2}{g}, \quad (1.10)$$

where  $\alpha$  is an arbitrary number. With (1.10) we have come close to obtaining our earlier result (1.7) and have done so without solving a differential equation or using



calculus to find the maximum value. Based on this rather minimal effort we can make the following observations:

- If the initial velocity is increased by a factor of 2, then the maximum height will increase by a factor of 4. This observation offers an easy method for experimentally checking on whether the original modeling assumptions are correct.
- The constant  $\alpha$  can be determined by running one experiment. Namely, for a given initial velocity  $v_0 = \bar{v}_0$  we measure the maximum height  $x_M = \bar{x}_M$ . With these known values,  $\alpha = g\bar{x}_M/\bar{v}_0^2$ . Once this is done, the formula in (1.10) can be used to determine  $x_M$  for any  $v_0$ .
- The maximum height does not depend on the mass of the object. This is not a surprise since the differential equation (1.4) and the initial conditions (1.2) and (1.3) do not depend on the mass.

The steps we have used are the basis for the method of dimensional reduction, where an expression is simplified based on the fundamental dimensions of the quantities involved. Given how easy it was to obtain (1.10) the method is very attractive as an analysis tool. It does have limitations and one is that we do not know the value of the number  $\alpha$ . It also requires us to be able to identify at the beginning what parameters are needed. The importance of this and how this relies on understanding the physical laws underlying the problem will be discussed later.

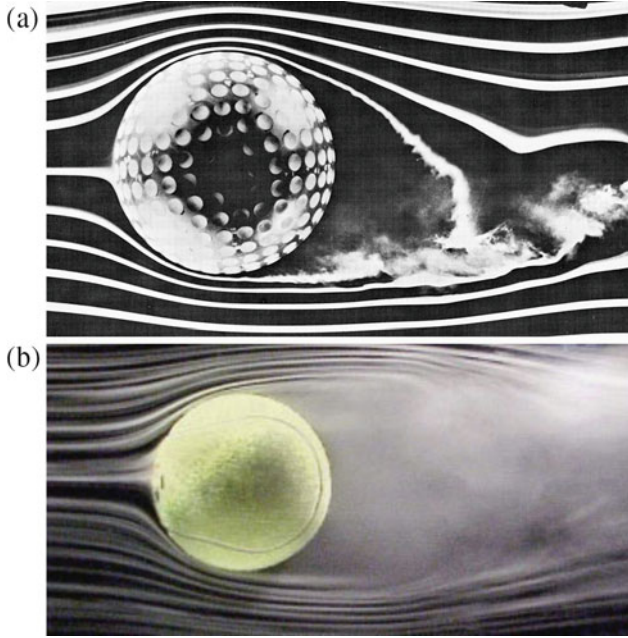
The purpose of the above example is to introduce the idea of dimensional reduction. What it does not show is how to handle problems with several parameters and this is the purpose of the next two examples.

### 1.2.2 Drag on a Sphere

In the design of automobiles, racing bicycles, and aircraft there is an overall objective to keep the drag on the object as small as possible. It is interesting to see what insight dimensional analysis might provide in such a situation, but since we are beginners it will be assumed the object is very simple and is a sphere (see Fig. 1.2). The modeling assumption that is made is that the drag force  $D_F$  on the sphere depends on the radius  $R$  of the sphere, the (positive) velocity  $v$  of the sphere, the density  $\rho$  of the air, and the dynamic viscosity  $\mu$  of the air. The latter is a measure of the resistance force of the air to motion and we will investigate this in Chap. 8. For the moment all we need is its fundamental dimensions and these are given in Table 1.1. In mathematical terms the modeling assumption is

$$D_F = f(R, v, \rho, \mu), \quad (1.11)$$

and we want to use dimensional reduction to find a simplified version of this expression. As will become evident in the derivation, this requires two steps.



**Fig. 1.2** Air flow around an object can be visualized using smoke. The flow around a golf ball is shown in (a) (Brown (1971)) and around a tennis ball in (b) (Bluck (2000)). In both cases the air is moving from left to right

### Find the General Product Solution

Similar to the last example, the first question is whether we can find numbers  $a$ ,  $b$ ,  $c$ , and  $d$ , so that

$$[D_F] = [R^a v^b \rho^c \mu^d]. \quad (1.12)$$

Expressing these using fundamental dimensions yields

$$\begin{aligned} M L T^{-2} &= L^a (L/T)^b (M/L^3)^c (M/LT)^d \\ &= L^{a+b-3c-d} T^{-b-d} M^{c+d}. \end{aligned}$$

As before, we equate the respective exponents and conclude

$$\begin{aligned} L : & a + b - 3c - d = 1, \\ T : & -b - d = -2, \\ M : & c + d = 1. \end{aligned} \quad (1.13)$$

We have four unknowns and three equations, so it is anticipated that in solving the above system of equations one of the constants will be undetermined. From the  $T$  equation we have  $b = 2 - d$ , and from the  $M$  equation  $c = 1 - d$ . The  $L$  equation then gives us  $a = 2 - d$ . With these solutions, and based on our assumption in (1.12), we have that

$$\begin{aligned} D_F &= \alpha R^{2-d} v^{2-d} \rho^{1-d} \mu^d \\ &= \alpha \rho R^2 v^2 \left( \frac{\mu}{R v \rho} \right)^d, \end{aligned}$$

where  $d$  and  $\alpha$  are arbitrary numbers. This can be written as

$$D_F = \alpha \rho R^2 v^2 \Pi^d, \quad (1.14)$$

where

$$\Pi = \frac{\mu}{R v \rho}. \quad (1.15)$$

This is the *general product solution* for how  $D_F$  depends on the given variables. The quantity  $\Pi$  is dimensionless, and it is an example of what is known as a dimensionless product. Calling it a product is a bit misleading as  $\Pi$  involves both multiplications and divisions. Some avoid this by calling it a dimensionless group. We will use both expressions in this book.

### Determine the General Solution

The formula for  $D_F$  in (1.14) is not the final answer. The conclusion that is derived from (1.14) is that the general solution is not an arbitrary power of  $\Pi$ , but it is an arbitrary function of  $\Pi$ . Mathematically, the conclusion is that the *general solution* can be written as

$$D_F = \rho R^2 v^2 F(\Pi), \quad (1.16)$$

where  $F$  is an arbitrary function of the dimensionless product  $\Pi$ . Note that because  $F$  is arbitrary, it is not necessary to include the multiplicative number  $\alpha$  that appears in (1.14).

To explain how (1.16) is derived from the general product solution (1.14), suppose you are given two sets of values for  $(\alpha, d)$ , say  $(\alpha_1, d_1)$  and  $(\alpha_2, d_2)$ . In this case, their sum

$$\begin{aligned} D_F &= \alpha_1 \rho R^2 v^2 \Pi^{d_1} + \alpha_2 \rho R^2 v^2 \Pi^{d_2} \\ &= \rho R^2 v^2 (\alpha_1 \Pi^{d_1} + \alpha_2 \Pi^{d_2}) \end{aligned}$$

is also a solution. This observation is not limited to just two sets of values, and, in fact, it holds for an arbitrary number. In other words,

$$D_F = \rho R^2 v^2 (\alpha_1 \Pi^{d_1} + \alpha_2 \Pi^{d_2} + \alpha_3 \Pi^{d_3} + \dots) \quad (1.17)$$

is a solution, where  $d_1, d_2, d_3, \dots$  are arbitrary numbers as are the coefficients  $\alpha_1, \alpha_2, \alpha_3, \dots$ . To express this in a more compact form, note that the expression within the parentheses in (1.17) is simply a function of  $\Pi$ . This is the reason for the  $F(\Pi)$  that appears in (1.16).

With the general solution in (1.16), we have used dimensional analysis to reduce the original assumption in (1.11), which involves an unknown of four variables, down to an unknown function of one variable. Although this is a significant improvement, the result is perhaps not as satisfying as the one obtained for the projectile example, given in (1.10), because we have not been able to determine  $F$ . However, there are various ways to address this issue, and some of them will be considered below.

### Representation of Solution

Now that the derivation is complete a few comments are in order. First, it is possible for two people to go through the above steps and come to what looks to be very different conclusions. For example, the general solution can also be written as

$$D_F = \frac{\mu^2}{\rho} H(\Pi), \quad (1.18)$$

where  $H$  is an arbitrary function of  $\Pi$ . The proof that this is equivalent to (1.16) comes from the requirement that the two expressions must produce the same result. In other words, it is required that

$$\frac{\mu^2}{\rho} H(\Pi) = \rho R^2 v^2 F(\Pi).$$

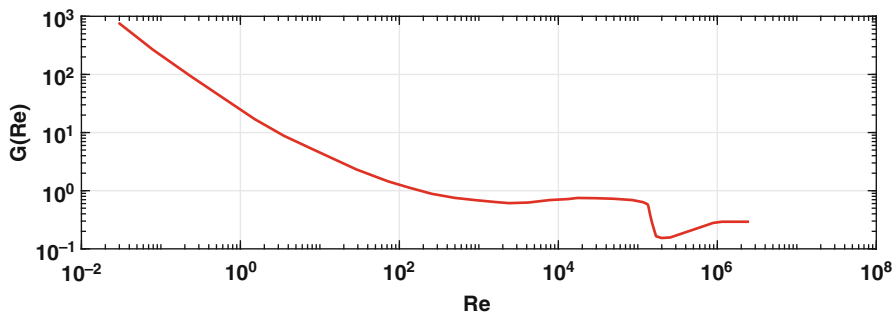
Solving this for  $H$  yields

$$H(\Pi) = \frac{1}{\Pi^2} F(\Pi).$$

The fact that the right-hand side of the above equation only depends on  $\Pi$  shows that (1.18) is equivalent to (1.16). As an example, if  $F(\Pi) = \Pi$  in (1.16), then  $H(\Pi) = 1/\Pi$  in (1.18).

Another representation for the general solution is

$$D_F = \rho R^2 v^2 G(Re), \quad (1.19)$$



**Fig. 1.3** The measured values of the function  $G(Re)$  that arises in the formula for the drag on a sphere, as given in (1.19)

where

$$Re = \frac{Rv\rho}{\mu}, \quad (1.20)$$

and  $G$  is an arbitrary function of  $Re$ . In fluid dynamics, the dimensionless product  $Re$  is known as the Reynolds number. To transform between the representation in (1.19), and the one in (1.16), note  $Re = 1/\Pi$ . From the requirement

$$\rho R^2 v^2 G(Re) = \rho R^2 v^2 F(\Pi),$$

we obtain

$$G(Re) = F(1/Re).$$

Because of its importance in fluids,  $G$  has been measured for a wide range of Reynolds numbers, producing the curve shown in Fig. 1.3. For those who have taken a course in fluid dynamics, the data in Fig. 1.3 are usually reported for what is called the drag coefficient  $C_D$  of a sphere. The two functions  $G$  and  $C_D$  are related through the equation  $G = \frac{\pi}{2} C_D$ .

The reason for the different representations is that there are four unknowns in (1.12) yet only three equations. This means one of the unknowns is used in the general solution and, as expressed in (1.14), we used  $d$ . If you were to use one of the others, then a different looking, but mathematically equivalent, expression would be obtained. The fact that there are multiple ways to express the solution can be used to advantage. For example, if one is interested in the value of  $D_F$  for small values of the velocity, then (1.19) would be a bit easier to use. The reason is that to investigate the case of small  $v$  it is somewhat easier to determine what happens to  $G$  for  $Re$  near zero than to expand  $F$  for large values of  $\Pi$ . For the same reason, (1.16) is easier to work with for studying large velocities. One last comment to make is that even

though there are choices on the form of the general solution, they all have exactly the same number of dimensionless products.

### Determining $F$

A more challenging question concerns how to determine the function  $F$  in (1.16). The mathematical approach would be to solve the equations for fluid flow around a sphere and from this find  $F$ . This is an intriguing idea but a difficult one since the equations are nonlinear partial differential equations (see Sect. 8.11). There is, however, another more applied approach that makes direct use of (1.16). Specifically, a sequence of experiments is run to measure  $F(r)$  for  $0 < r < \infty$ . To do this, a sphere with a given radius  $R_0$ , and a fluid with known density  $\rho_0$  and viscosity  $\mu_0$ , are selected. In this case (1.16) can be written as

$$F(r) = \frac{\gamma D_F}{v^2}, \quad (1.21)$$

where  $\gamma = 1/(\rho_0 R_0^2)$  is known and fixed. The experiment consists of taking various values of  $v$  and then measuring the resulting drag force  $D_F$  on the sphere. To illustrate, suppose our choice for the sphere and fluid give  $R = 1$ ,  $\rho_0 = 2$ , and  $\mu_0 = 3$ . Also, suppose that running the experiment using  $v = 4$  produces a measured drag of  $D_F = 5$ . In this case  $r = \mu_0/(R_0 v \rho_0) = 3/8$  and  $\gamma D_F/v^2 = 5/32$ . Our conclusion is therefore that  $F(3/8) = 5/32$ . In this way, picking a wide range of  $v$  values we will be able to determine the values for the function  $F(r)$ . This approach is used extensively in the real world and the example we are considering has been a particular favorite for study. The data determined from such experiments are shown in Fig. 1.3.

A number of conclusions can be drawn from Fig. 1.3. For example, there is a range of  $Re$  values where  $G$  is approximately constant. Specifically, if  $10^3 < Re < 10^5$ , then  $G \approx 0.7$ . This is the reason why in the fluid dynamics literature you will occasionally see the statement that the drag coefficient  $C_D = \frac{2}{\pi} G$  for a sphere has a constant value of approximately 0.44. For other  $Re$  values, however,  $G$  is not constant. Of particular interest is the dependence of  $G$  for small values of  $Re$ . This corresponds to velocities  $v$  that are very small, what is known as Stokes flow. The data in Fig. 1.3 show that  $G$  decreases linearly with  $Re$  in this region. Given that this is a log-log plot, then this means that  $\log(G) = a - b \log(Re)$ , or equivalently,  $G = \alpha/Re^b$  where  $\alpha = 10^a$ . Curve fitting this function to the data in Fig. 1.3 it is found that  $\alpha \approx 17.6$  and  $\beta \approx 1.07$ . These are close to the exact values of  $\alpha = 6\pi$  and  $\beta = 1$ , which are obtained by solving the equations of motion for Stokes flow. Inserting these values into (1.19), the conclusion is that the drag on the sphere for small values of the Reynolds number is

$$D_F \approx 6\pi \mu R v. \quad (1.22)$$

This is known as Stokes formula for the drag on a sphere, and we will have use for it in Chap. 4 when studying diffusion.

## Scale Models

Why all the work to find  $F$ ? Well, knowing this function allows for the use of scale model testing. To explain, suppose it is required to determine the drag on a sphere with radius  $R_f$  for a given velocity  $v_f$  when the fluid has density  $\rho_f$  and viscosity  $\mu_f$ . Based on (1.16) we have  $D_F = \rho_f R_f^2 v_f^2 F(\Pi_f)$ , where

$$\Pi_f = \frac{\mu_f}{R_f v_f \rho_f}. \quad (1.23)$$

Consequently, we can determine  $D_F$  if we know the value of  $F$  at  $\Pi_f$ . Also, suppose that this cannot be measured directly as  $R_f$  is large and our experimental equipment can only handle small spheres. We can still measure  $F(\Pi_f)$  using a small value of  $R$  if we change one or more of the parameters in such a way that the value of  $\Pi_f$  does not change. If  $R_m$ ,  $\mu_m$ ,  $\rho_m$ , and  $v_m$  are the values used in the experiment, then we want to select them so that

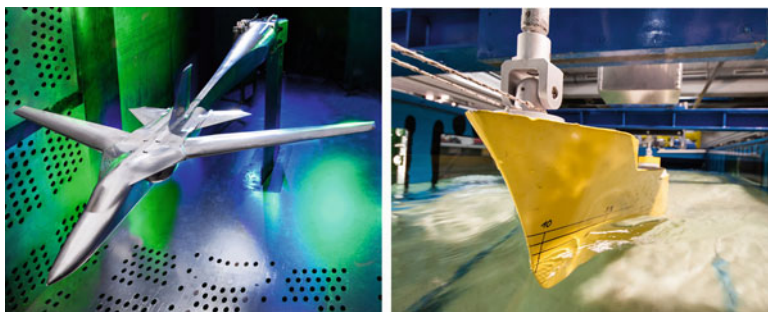
$$\frac{\mu_m}{R_m v_m \rho_m} = \frac{\mu_f}{R_f v_f \rho_f}, \quad (1.24)$$

or equivalently

$$v_m = \frac{\mu_m R_f \rho_f}{\mu_f R_m \rho_m} v_f. \quad (1.25)$$

This equation relates the values for the full-scale ball (subscript  $f$ ) to those for the model used in the experiment (subscript  $m$ ). As an example, suppose we are interested in the drag on a very large sphere, say  $R_f = 100\text{ m}$ , but our equipment can only handle smaller values, say  $R_m = 2\text{ m}$ . If the fluid for the two cases is the same, so  $\rho_m = \rho_f$  and  $\mu_m = \mu_f$ , then according to (1.25), in our experiment we should take  $v_m = 50v_f$ . If the experimental apparatus is unable to generate velocities 50 times the value of  $v_f$ , then it would be necessary to use a different fluid to reduce this multiplicative factor.

The result in the above example is the basis of scale model testing used in wind tunnels and towing tanks (see Fig. 1.4). Usually these tests involve more than just keeping one dimensionless product constant as we did in (1.24). Moreover, it is evident in Fig. 1.4 that the models look like the originals, they are just smaller. This is the basis of geometric similarity, where the lengths of the model are all a fraction of the original. Other scalings are sometimes used and the most common are kinematic similarity, where velocities are scaled, and dynamic similarity, where forces are scaled.



**Fig. 1.4** Dimensional analysis is used in the development of scale model testing. On the left is a test in NASA's Glenn Research Center supersonic wind tunnel (NASA, 2018), and on the right a towing tank test of a model for a ship hull (el Moctar, 2018)

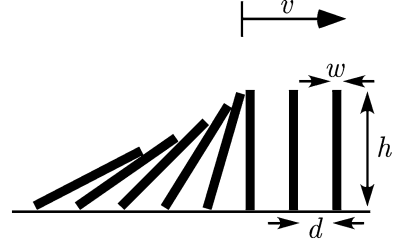
## Endnotes

One question that has not been considered so far is, how do you know to assume that the drag force depends on the radius, velocity, density, and dynamic viscosity? The assumption comes from knowing the laws of fluid dynamics, and identifying the principal terms that contribute to the drag. For the most part, in this chapter the assumptions will be stated explicitly, as they were in this example. Later in the text, after the basic physical laws are developed, it will be possible to construct the assumptions directly. However, one important observation can be made, and that is the parameters used in the assumption should be independent. For example, even though the drag on a sphere likely depends on the surface area and volume of the sphere it is not necessary to include them in the list. The reason is that it is already assumed that  $D_F$  depends on the radius  $R$  and both the surface area and volume are determined using  $R$ .

The problem of determining the drag on a sphere is one of the oldest in fluid dynamics. Given that the subject is well over 150 years old, you would think that whatever useful information can be derived from this particular problem was figured out long ago. Well, apparently not, as research papers still appear regularly on this topic. A number of them come from the sports industry, where there is interest in the drag on soccer balls (Asai et al. 2007), golf balls (Smits and Ogg 2004), tennis balls (Goodwill et al. 2004), as well as nonspherical-shaped balls (Mehta 1985). Others have worked on how to improve the data in Fig. 1.3, and an example is the use of a magnetic suspension system to hold the sphere (Sawada and Kunimasu 2004). A more novel idea is to drop different types of spheres down a deep mine shaft, and then use the splash time as a means to determine the drag coefficient (Maroto et al. 2005). The point here is that even the most studied problems in science and engineering still have interesting questions that remain unanswered.



**Fig. 1.5** Schematic of toppling dominoes, creating a wave that propagates with velocity  $v$



### 1.2.3 Toppling Dominoes

Domino toppling refers to the art of setting up dominoes, and then knocking them down. The current world record for this is about 4,500,000 dominoes for a team, and 320,000 for an individual. One of the more interesting aspects of this activity is that as the dominoes fall it appears as if a wave is propagating along the line of dominoes. The objective of this example is to examine what dimensional analysis might be able to tell us about the velocity of this wave. A schematic of the situation is shown in Fig. 1.5. The assumption is that the velocity  $v$  depends on the spacing  $d$ , height  $h$ , thickness  $w$ , and the gravitational acceleration constant  $g$ . Therefore, the modeling assumption is  $v = f(d, h, w, g)$  and we want to use dimensional reduction to find a simplified version of this expression. As usual, the first step is to find numbers  $a$ ,  $b$ ,  $c$ , and  $e$  so that

$$[v] = [d^a h^b w^c g^e].$$

Expressing these using fundamental dimensions yields

$$\begin{aligned} LT^{-1} &= L^a L^b L^c (L/T^2)^e \\ &= L^{a+b+c+e} T^{-2e}. \end{aligned}$$

Equating the respective terms we obtain

$$\begin{aligned} L : a + b + c + e &= 1, \\ T : -2e &= -1. \end{aligned}$$

Solving these two equations gives us that  $e = \frac{1}{2}$  and  $b = \frac{1}{2} - a - c$ . With this we have that

$$\begin{aligned} v &= \alpha d^a h^{1/2-a-c} w^c g^{1/2} \\ &= \alpha \sqrt{hg} \left(\frac{d}{h}\right)^a \left(\frac{w}{h}\right)^c \\ &= \alpha \sqrt{hg} \Pi_1^a \Pi_2^c, \end{aligned} \tag{1.26}$$

where  $\alpha$  is an arbitrary number, and the two dimensionless products are

$$\Pi_1 = \frac{d}{h},$$

$$\Pi_2 = \frac{w}{h}.$$

The expression in (1.26) is the general product solution. Therefore, the general solution for how the velocity depends on the given parameters is

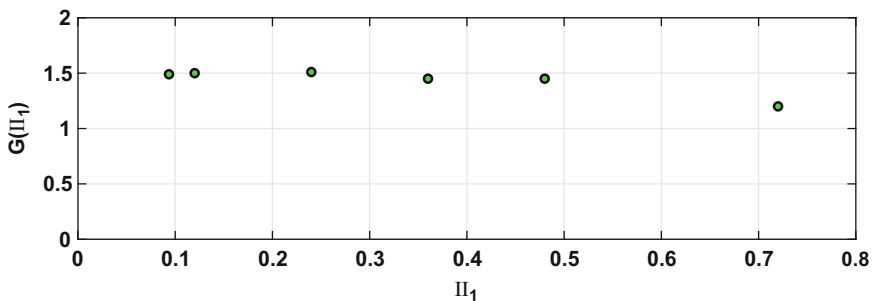
$$v = \sqrt{hg} F(\Pi_1, \Pi_2), \quad (1.27)$$

where  $F$  is an arbitrary function of the two dimensionless products. An explanation of how (1.27) follows from (1.26) is very similar to the method used to derive (1.16) from (1.15). This is discussed in Exercise 1.26.

Dimensional analysis has been able to reduce the original assumption involving a function of four-dimensional parameters down to one involving two dimensionless products. This example is also informative as it demonstrates how to obtain the general solution when more than one dimensionless product is involved. The question remains, however, if this really applies to toppling dominoes. It does, but in using this formula it is usually assumed the dominoes are very thin, or more specifically that  $w \ll h$ . This means that it is possible to assume  $\Pi_2 = 0$ , and (1.27) simplifies to

$$v = \sqrt{hg} G(\Pi_1), \quad (1.28)$$

where  $G$  is an arbitrary function. Some effort has been made to measure  $G$ , and the measurements for a particular type of domino are given in Fig. 1.6. It is seen that for smaller values of  $\Pi_1$ ,  $G \approx 1.5$ . Therefore, as an approximation we conclude that the speed at which dominoes topple is  $v \approx 1.5\sqrt{hg}$ . A typical domino has  $h = 5$  cm, which results in a velocity of  $v \approx 1$  m/s. To obtain a more explicit formula for



**Fig. 1.6** Data for toppling dominoes (Stronge and Shu 1988). In these experiments,  $w = 0.12h$ , so the thin domino approximation is appropriate

$G$ , however, requires the solution of a challenging mathematical problem, and an expanded discussion of this can be found in Efthimiou and Johnson (2007).

### 1.2.4 Endnotes

Based on the previous examples, the benefits of using dimensional reduction are apparent. However, a word of caution is needed here as the method gives the impression that it is possible to derive useful information without getting involved with the laws of physics or potentially difficult mathematical problems. One consequence of this is that the method is used to comment on situations and phenomena that are simply inappropriate (e.g., to study psychoacoustic behavior). The method relies heavily on knowing the fundamental laws for the problem under study, and without this whatever conclusions made using dimensional reduction are limited. For example, we earlier considered the drag on a sphere and in the formulation of the problem we assumed that the drag depends on the dynamic viscosity. Without knowing the equations of motion for fluids it would not have been possible to know that this term needed to be included or what units it might have. By not including it we would have concluded that  $d = 0$  in (1.14) and instead of (1.16) we would have  $D_F = \alpha \rho R^2 v^2$  where  $\alpha$  is a constant. In Fig. 1.3 it does appear that  $D_F$  is approximately independent of  $Re$  when  $10^3 < Re < 10^5$ . However, outside of this interval,  $D_F$  is strongly dependent on  $Re$ , and this means ignoring the viscosity would be a mistake. Another example illustrating the need to know the underlying physical laws arises in the projectile problem when we included the gravitational constant. Again, this term is essential and without some understanding of Newtonian mechanics it would be missed completely. The point here is that dimensional reduction can be a very effective method for simplifying complex relationships, but it is based on knowing what the underlying laws are that govern the systems being studied.

## 1.3 Theoretical Foundation

The theoretical foundation for dimensional reduction is contained in the Buckingham Pi Theorem. To derive this result, assume we have a physical quantity  $q$  that depends on physical parameters or variables  $p_1, p_2, \dots, p_n$ . In this context, the word physical means that the quantity is measurable. Each can be expressed in fundamental dimensions and we will assume that the  $L, T, M$  system is sufficient for this task. In this case we can write

$$[q] = L^{\ell_0} T^{t_0} M^{m_0}, \quad (1.29)$$

and

$$\llbracket p_i \rrbracket = L^{\ell_i} T^{t_i} M^{m_i}. \quad (1.30)$$

Our modeling assumption is that  $q = f(p_1, p_2, \dots, p_n)$ , and that this is *dimensionally homogeneous*. This means, as explained earlier, that this formula is independent of the system of units used to measure  $q$  or the  $p_i$ 's.

To dimensionally reduce the equation, we will determine if there are numbers  $a_1, a_2, \dots, a_n$  so that

$$\llbracket q \rrbracket = \llbracket p_1^{a_1} p_2^{a_2} \cdots p_n^{a_n} \rrbracket. \quad (1.31)$$

Introducing (1.29) and (1.30) into the above expression, and then equating exponents, we obtain the equations

$$\begin{aligned} L : \quad & \ell_1 a_1 + \ell_2 a_2 + \cdots + \ell_n a_n = \ell_0, \\ T : \quad & t_1 a_1 + t_2 a_2 + \cdots + t_n a_n = t_0, \\ M : \quad & m_1 a_1 + m_2 a_2 + \cdots + m_n a_n = m_0. \end{aligned}$$

This can be expressed in matrix form as

$$\mathbf{A}\mathbf{a} = \mathbf{b}, \quad (1.32)$$

where

$$\mathbf{A} = \begin{pmatrix} \ell_1 & \ell_2 & \cdots & \ell_n \\ t_1 & t_2 & \cdots & t_n \\ m_1 & m_2 & \cdots & m_n \end{pmatrix}, \quad (1.33)$$

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \ell_0 \\ t_0 \\ m_0 \end{pmatrix}. \quad (1.34)$$

The matrix  $\mathbf{A}$  is known as the dimension matrix. As expressed in (1.33) it is  $3 \times n$  but if we were to have used  $L, T, M, \theta$  as the fundamental system, then it would be  $4 \times n$ . In other words, the number of rows in the dimension matrix equals the number of fundamental units needed, and the number of columns equals the number of parameters that  $q$  is assumed to depend on.

With (1.32) we have transformed the dimensional reduction question into a linear algebra problem. To determine the consequences of this we first consider the situation that (1.32) has no solution. In this case the assumption that  $q$  depends on

$p_1, p_2, \dots, p_n$  is incomplete and additional parameters are needed. This situation motivates the following definition.

**Dimensionally Complete.** *The set  $p_1, p_2, \dots, p_n$  is dimensionally complete for  $q$  if it is possible to combine the  $p_i$ 's to produce a quantity with the same dimension as  $q$ . If it is not possible, the set is dimensionally incomplete for  $q$ .*

From this point on we will assume the  $p_i$ 's are complete so there is at least one solution of (1.32).

To write down the general solution we consider the associated homogeneous equation, namely  $\mathbf{A}\mathbf{a} = \mathbf{0}$ . The set of solutions of this equation form a subspace  $N(\mathbf{A})$ , known as the null space of  $\mathbf{A}$ . Letting  $k$  be the dimension of this subspace, then the general solution of  $\mathbf{A}\mathbf{a} = \mathbf{0}$  can be written as  $\mathbf{a} = \gamma_1\mathbf{a}_1 + \gamma_2\mathbf{a}_2 + \dots + \gamma_k\mathbf{a}_k$ , where  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k$  is a basis for  $N(\mathbf{A})$  and  $\gamma_1, \gamma_2, \dots, \gamma_k$  are arbitrary. It is understood here that if  $k = 0$ , then  $\mathbf{a} = \mathbf{0}$ . With this, the general solution of (1.32) can be written as

$$\mathbf{a} = \mathbf{a}_p + \gamma_1\mathbf{a}_1 + \gamma_2\mathbf{a}_2 + \dots + \gamma_k\mathbf{a}_k, \quad (1.35)$$

where  $\mathbf{a}_p$  is any vector that satisfies (1.32) and  $\gamma_1, \gamma_2, \dots, \gamma_k$  are arbitrary numbers.

*Example (Drag on a Sphere)* To connect the above discussion with what we did earlier, consider the drag on a sphere example. Writing (1.13) in matrix form we obtain

$$\begin{pmatrix} 1 & 1 & -3 & -1 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}.$$

This is the matrix equation (1.32) for this particular example. Putting this in augmented form, and row reducing, yields the following

$$\left( \begin{array}{cccc|c} 1 & 1 & -3 & -1 & 1 \\ 0 & -1 & 0 & -1 & -2 \\ 0 & 0 & 1 & 1 & 1 \end{array} \right) \rightarrow \left( \begin{array}{cccc|c} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 1 \end{array} \right).$$

From this we conclude that  $a = 2 - d$ ,  $b = 2 - d$ , and  $c = 1 - d$ . To be consistent with the notation in (1.35), set  $d = \gamma$ , so the solution is

$$\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 1 \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} -1 \\ -1 \\ -1 \\ 1 \end{pmatrix},$$

where  $\gamma$  is arbitrary. Comparing this with (1.35) we have that  $k = 1$ ,

$$\mathbf{a}_p = \begin{pmatrix} 2 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad \mathbf{a}_1 = \begin{pmatrix} -1 \\ -1 \\ -1 \\ 1 \end{pmatrix}. \quad \blacksquare$$

It is now time to take our linear algebra conclusions and apply them to the dimensional reduction problem. Just as the appearance of  $d$  in (1.14) translated into the appearance of a dimensionless product in the general solution given in (1.16), each of the  $\gamma_i$ 's in (1.35) gives rise to a dimensionless product in the general solution for the problem we are currently studying. To be specific, writing the  $i$ th basis vector  $\mathbf{a}_i$  in component form as

$$\mathbf{a}_i = \begin{pmatrix} \alpha \\ \beta \\ \vdots \\ \gamma \end{pmatrix}, \quad (1.36)$$

then the corresponding dimensionless product is

$$\Pi_i = p_1^\alpha p_2^\beta \cdots p_n^\gamma. \quad (1.37)$$

Moreover, because the  $\mathbf{a}_i$ 's are independent vectors, the dimensionless products  $\Pi_1, \Pi_2, \dots, \Pi_k$  are independent.

As for the particular solution  $\mathbf{a}_p$  in (1.35), assuming it has components

$$\mathbf{a}_p = \begin{pmatrix} a \\ b \\ \vdots \\ c \end{pmatrix}, \quad (1.38)$$

then the quantity

$$Q = p_1^a p_2^b \cdots p_n^c \quad (1.39)$$

has the same dimensions as  $q$ .

Based on the conclusions of the previous two paragraphs, the general product solution is  $q = \alpha Q \Pi_1^{\kappa_1} \Pi_2^{\kappa_2} \cdots \Pi_k^{\kappa_k}$ , where  $\alpha, \kappa_1, \kappa_2, \dots, \kappa_k$  are arbitrary constants. The form of the resulting general solution is given in the following theorem.

**Buckingham Pi Theorem.** *Assuming the formula  $q = f(p_1, p_2, \dots, p_n)$  is dimensionally homogeneous and dimensionally complete, then it is possible to reduce it to one of the form  $q = QF(\Pi_1, \Pi_2, \dots, \Pi_k)$ , where  $\Pi_1, \Pi_2, \dots, \Pi_k$  are independent dimensionless products of  $p_1, p_2, \dots, p_n$ . The quantity  $Q$  is a dimensional product of  $p_1, p_2, \dots, p_n$  with the same dimensions as  $q$ .*

According to this theorem, the original formula for  $q$  can be reduced from a function of  $n$  variables down to one with  $k$ . The value of  $k$ , which equals the nullity of the dimension matrix, ranges from 0 to  $n - 1$  depending on the given quantities  $p_1, p_2, \dots, p_n$ . If it happens that  $k = 0$ , then the function  $F$  reduces to a constant and the conclusion is that  $q = \alpha Q$ , where  $\alpha$  is an arbitrary number.

The importance of this theorem is that it establishes that the process used to reduce the drag on a sphere and toppling dominoes examples can be applied to more complex problems. It also provides insight into how the number of dimensionless products is determined. There are still, however, fundamental questions left unanswered. For example, those with a more mathematical bent might still be wondering if this result can really be true no matter how discontinuous the original function  $f$  might be. Others might be wondering if the fundamental units used here, particularly length and time, are really independent. This depth of inquiry, although quite interesting, is beyond the scope of this text. Those wishing to pursue further study of these and related topics should consult Penrose (2007) and Bluman and Anco (2002).

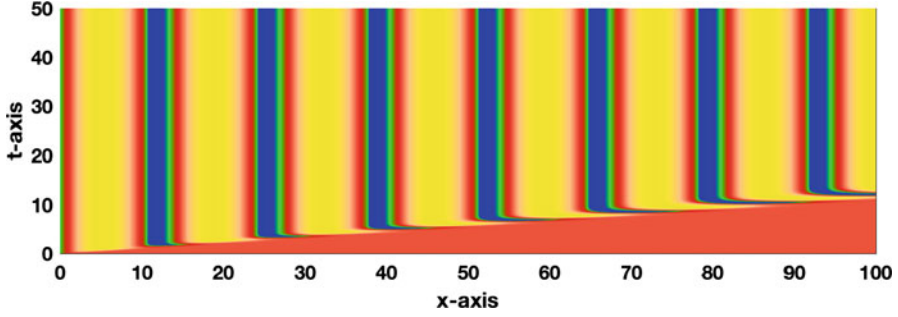
### 1.3.1 Pattern Formation

The mechanism responsible for the colorful patterns on seashells, butterfly wings, zebras, and the like has intrigued scientists for decades. An experiment that has been developed to study pattern formation involves pouring chemicals into one end of a long tube, and then watching what happens as they interact while moving along the tube. This apparatus is called a plug-flow reactor and the outcome of a mathematical model coming from such an experiment is shown in Fig. 1.7. It was found in these experiments that patterns appear only for certain pouring velocities  $v$ . According to what is known as the Lengyel-Epstein model, this velocity depends on the concentration  $U$  of the chemical used in the experiment, the rate  $k_2$  at which the chemicals interact, the diffusion coefficient  $D$  of the chemicals, and a parameter  $k_3$  that has the dimensions of concentration squared. The model is therefore assuming

$$v = f(U, k_2, D, k_3). \quad (1.40)$$

From Table 1.1 we have that  $\llbracket v \rrbracket = L/T$ ,  $\llbracket U \rrbracket = 1/L^3$ ,  $\llbracket D \rrbracket = L^2/T$ , and  $\llbracket k_3 \rrbracket = 1/L^6$ . Also, from the Lengyel-Epstein model one finds that  $\llbracket k_2 \rrbracket = L^3/T$ . Using dimensional reduction we require

$$\llbracket v \rrbracket = \llbracket U^a k_2^b D^c k_3^d \rrbracket. \quad (1.41)$$



**Fig. 1.7** Spatial pattern created from a model for a plug-flow reactor. The tube occupies the interval  $0 \leq x \leq 100$ , and starting at  $t = 0$  the chemicals are poured into the left end. As they flow along the tube a striped pattern develops

Expressing these using fundamental dimensions yields

$$\begin{aligned} LT^{-1} &= (L^{-3})^a (L^3 T^{-1})^b (L^2 T^{-1})^c (L^{-6})^d \\ &= L^{-3a+3b+2c-6d} T^{-b-c}. \end{aligned}$$

As before we equate the respective terms and conclude

$$\begin{aligned} L : -3a + 3b + 2c - 6d &= 1 \\ T : -b - c &= -1. \end{aligned}$$

These equations will enable us to express two of the unknowns in terms of the other two. There is no unique way to do this, and one choice yields  $b = -1 + 3a + 6d$  and  $c = 2 - 3a - 6d$ . From this it follows that the general product solution is

$$\begin{aligned} v &= \alpha U^a k_2^{3a+6d-1} D^{2-3a-6d} k_3^d \\ &= \alpha k_2^{-1} D^2 (U k_2^3 D^{-3})^a (k_2^6 D^{-6} k_3)^d. \end{aligned}$$

This can be rewritten as

$$v = \alpha k_2^{-1} D^2 \Pi_1^a \Pi_2^d, \quad (1.42)$$

where

$$\Pi_1 = \frac{U k_2^3}{D^3}, \quad (1.43)$$

and

$$\Pi_2 = \frac{k_2^6 k_3}{D^6}. \quad (1.44)$$



The dimensionless products  $\Pi_1$  and  $\Pi_2$  are independent, and this follows from the method used to derive these expressions. Independence is also evident from the observation that  $\Pi_1$  and  $\Pi_2$  do not involve exactly the same parameters. From this result it follows that the general form of the reduced equation is

$$v = k_2^{-1} D^2 F(\Pi_1, \Pi_2). \quad (1.45)$$

It is of interest to compare (1.45) with the exact formula obtained from solving the differential equations coming from the Lengyel-Epstein model. It is found that

$$v = \sqrt{k_2 D U} G(\beta), \quad (1.46)$$

where  $\beta = k_3/U^2$  and  $G$  is a rather complicated square root function (Bamforth et al. 2000). This result appears to differ from (1.45). To investigate this, note that  $\beta = \Pi_2/\Pi_1^2$ . Equating (1.45) and (1.46) it follows that

$$\begin{aligned} F(\Pi_1, \Pi_2) &= \frac{k_2^{3/2} U^{1/2}}{D^{3/2}} G(\beta) \\ &= \sqrt{\Pi_1} G(\Pi_2/\Pi_1^2). \end{aligned}$$

Because the right-hand side is a function of only  $\Pi_1$  and  $\Pi_2$ , then (1.45) does indeed reduce to the exact result (1.46).

Dimensional reduction has therefore successfully reduced the original unknown function of four variables in (1.40) down to one with only two variables. However, the procedure is not able to reduce the function down to one dimensionless variable, as given in (1.46). In this problem that level of reduction requires information only available from the mathematical problem coming from the Lengyel-Epstein model. An illustration of how this is done can be found in Exercises 1.29 and 1.30. It is worth noting that the method for deriving the Lengyel-Epstein model is explained in Sect. 4.6.3.

## 1.4 Similarity Variables

Dimensions can be used not only to reduce formulas but also to simplify complex mathematical problems. The degree of simplification depends on the parameters, and variables, in the problem. One of the more well-known examples is the problem of finding the mass density  $u(x, t)$ . In this case the density satisfies the diffusion equation

$$D \frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t}, \quad \text{for } \begin{cases} 0 < x < \infty, \\ 0 < t, \end{cases} \quad (1.47)$$

where the boundary conditions are

$$u|_{x=0} = u_0, \quad u|_{x \rightarrow \infty} = 0, \quad (1.48)$$

and the initial condition is

$$u|_{t=0} = 0. \quad (1.49)$$

It is assumed that  $D$  is positive and  $u_0$  is nonzero.

The constant  $D$  is called the diffusion coefficient, and its dimensions can be determined from the terms in the differential equation. To do this, it is useful to know the following facts, all of which come directly from the definition of the derivative.

- Given  $f(t)$ , then

$$\left[ \frac{df}{dt} \right] = \frac{[f]}{[t]}, \quad \text{and} \quad \left[ \frac{d^2 f}{dt^2} \right] = \frac{[f]}{[t]^2}. \quad (1.50)$$

- Given  $u(x, t)$ , then

$$\left[ \frac{\partial u}{\partial t} \right] = \frac{[u]}{[t]}, \quad \left[ \frac{\partial u}{\partial x} \right] = \frac{[u]}{[x]} \quad \text{and} \quad \left[ \frac{\partial^2 u}{\partial t \partial x} \right] = \frac{[u]}{[t][x]}. \quad (1.51)$$

Some of the consequences, and extensions, of the above formulas are explored in Exercise 1.27.

Now, the dimensions of the left and right sides of (1.47) must be the same, and this means  $[Du_{xx}] = [u_t]$ . Because  $[u] = M/L^3$ , then  $[u_{xx}] = [u]/L^2 = M/L^5$  and  $[u_t] = [u]/T = M/(TL^3)$ . From this we have  $[D]M/L^5 = M/(TL^3)$ , and therefore  $[D] = L^2/T$ . In a similar manner, in boundary condition (1.48),  $[u_0] = [u] = M/L^3$ . As a final comment, the physical assumptions underlying the derivation of (1.47) are the subject of Chap. 4. In fact, the solution we are about to derive is needed in Sect. 4.6.2 to solve the diffusion equation.

### 1.4.1 Dimensional Reduction

The conventional method for solving the diffusion equation on a semi-infinite spatial interval is to use an integral transform, and this will be considered in Chap. 4. It is also possible to find  $u$  using dimensional reduction. The approach is based on the observation that the only dimensional variables, and parameters, appearing in the problem are  $u$ ,  $u_0$ ,  $D$ ,  $x$ , and  $t$ . In other words, it must be true that  $u = f(x, t, D, u_0)$ . With this we have the framework for dimensional reduction, and the question is whether we can find numbers  $a$ ,  $b$ ,  $c$ ,  $d$  so that

$$[[u]] = [[x^a t^b D^c (u_0)^d]]. \quad (1.52)$$

Using fundamental dimensions,

$$\begin{aligned} ML^{-3} &= L^a T^b (L^2/T)^c (M/L^3)^d \\ &= L^{a+2c-3d} T^{b-c} M^d, \end{aligned}$$

and then equating the respective terms gives us

$$\begin{aligned} L : a + 2c - 3d &= -3, \\ T : b - c &= 0, \\ M : d &= 1. \end{aligned} \quad (1.53)$$

The solution of the above system can be written as  $d = 1$  and  $b = c = -a/2$ . Given the assumption in (1.52), we conclude that the general product solution is

$$u = \alpha u_0 \left( \frac{x}{\sqrt{Dt}} \right)^a.$$

The general solution therefore has the form

$$u = u_0 F(\eta), \quad (1.54)$$

where

$$\eta = \frac{x}{\sqrt{Dt}}. \quad (1.55)$$

In this case,  $\eta$  is called a *similarity variable* as it is a dimensionless product that involves the independent variables in the problem.

When working out the drag on a sphere example, we discussed how it is possible to derive different representations of the solution. For the current example, when solving (1.53), instead of writing  $b = c = -a/2$ , we could just as well state that  $a = -2b$  and  $c = b$ . In this case (1.54) is replaced with  $u = u_0 G(\xi)$  where  $\xi = Dt/x^2$ . Although the two representations are equivalent, in the sense that one can be transformed into the other, it does make a difference which one is used when deriving a similarity solution. The reason is that (1.47) requires two derivatives with respect to  $x$ , and the resulting formulas are simpler if the similarity variable is a linear function of  $x$ . If you would like a hands on example of why this is true, try working out the steps below using the representation  $u = u_0 G(\xi)$  instead of (1.54).

### 1.4.2 Similarity Solution

Up to this point we have been using a routine dimensional reduction argument. Our result, given in (1.54), is interesting as it states that the solution has a very specific dependence on the independent variables  $x$  and  $t$ . Namely,  $u$  can be written as a function of a single intermediate variable  $\eta$ . To determine  $F$  we substitute (1.54) back into the problem and find what equation  $F$  satisfies. With this in mind note, using the chain rule,

$$\begin{aligned}\frac{\partial u}{\partial t} &= u_0 F'(\eta) \frac{\partial \eta}{\partial t} \\ &= u_0 F'(\eta) \left( -\frac{x}{2D^{1/2}t^{3/2}} \right) \\ &= -u_0 F'(\eta) \frac{\eta}{2t}.\end{aligned}$$

In a similar manner one finds that

$$\frac{\partial^2 u}{\partial x^2} = u_0 F''(\eta) \frac{1}{Dt}.$$

Substituting these into (1.47) yields

$$F'' = -\frac{1}{2}\eta F'.$$

Also, since  $0 < x < \infty$ , and  $t > 0$ , then  $0 < \eta < \infty$ .

We must also transform the boundary and initial conditions.

$u|_{x=0} = u_0$ : Letting  $x = 0$  in (1.54) yields  $u_0 F(0) = u_0$ , and from this we conclude that  $F(0) = 1$ .

$u|_{x \rightarrow \infty} = 0$ : Letting  $x \rightarrow \infty$  in (1.54) yields  $u_0 F(\infty) = 0$ , and from this we conclude that  $F(\infty) = 0$ .

$u|_{t=0} = 0$ : Given that  $\eta = x/\sqrt{Dt}$ , this condition must be dealt with using a limit. Specifically, the requirement is that

$$\lim_{t \rightarrow 0^+} u_0 F\left(\frac{x}{\sqrt{Dt}}\right) = 0.$$

For  $0 < x < \infty$ , the above limit gives us that  $F(\infty) = 0$ . This is the same condition we derived for  $u(\infty, t) = 0$ .

To summarize the above reduction, we have shown that the original diffusion problem can be replaced with solving

$$F'' = -\frac{1}{2}\eta F', \quad \text{for } 0 < \eta < \infty, \quad (1.56)$$

where

$$F(0) = 1, \quad (1.57)$$

and

$$F(\infty) = 0. \quad (1.58)$$

With this, we have transformed a problem involving a partial differential equation (PDE) into one with an ordinary differential equation (ODE). As required, the resulting problem for  $F$  is only in terms of  $\eta$ . All of the original dimensional quantities, including the independent variables  $x$  and  $t$ , do not appear anywhere in the problem. This applies not just to the differential equation, but also to the boundary and initial conditions.

The reduced problem is simple enough that it is possible to solve for  $F$ . This can be done by letting  $G = F'$ , so (1.56) takes the form  $G' = -\frac{1}{2}\eta G$ . The general solution of this is  $G = \alpha \exp(-\eta^2/4)$ . Because  $F' = G$ , we conclude that the general solution is

$$F(\eta) = \beta + \alpha \int_0^\eta e^{-s^2/4} ds. \quad (1.59)$$

From (1.57) we have that  $\beta = 1$  and from (1.58) we get

$$1 + \alpha \int_0^\infty e^{-s^2/4} ds = 0. \quad (1.60)$$

Given that  $\int_0^\infty e^{-s^2/4} ds = \sqrt{\pi}$ , then

$$\begin{aligned} F(\eta) &= 1 - \frac{1}{\sqrt{\pi}} \int_0^\eta e^{-s^2/4} ds \\ &= 1 - \frac{2}{\sqrt{\pi}} \int_0^{\eta/2} e^{-r^2} dr. \end{aligned} \quad (1.61)$$

Expressions like this arise so often that they have given rise to a special function known as the *complementary error function*  $\text{erfc}(z)$ . This is defined as

$$\text{erfc}(z) \equiv 1 - \frac{2}{\sqrt{\pi}} \int_0^z e^{-r^2} dr. \quad (1.62)$$

Therefore, we have found that the solution of the diffusion problem is

$$u(x, t) = u_0 \text{erfc}\left(\frac{x}{2\sqrt{Dt}}\right). \quad (1.63)$$

As the above example demonstrates, using similarity variables and dimensional analysis provides a powerful tool for solving PDEs. It is, for example, one of the very few methods known that can be used to solve nonlinear PDEs. Its limitation is that the problem must have a specific form to work. To illustrate, if the spatial interval in the above diffusion problem is changed to one that is finite, so  $0 < x < \ell$ , then dimensional analysis will show that there are two independent similarity variables. This represents no improvement as we already know it is a function of two independent variables, so a reduction is not possible. In some cases it is possible to take advantage of particular properties of the solution so a similarity reduction is possible, and this is illustrated in Exercises 1.29 and 1.30. Those interested in pursuing this a bit more should consult Bluman et al. (2010) and Hydon (2000).

## 1.5 Nondimensionalization and Scaling

Another use we will have for dimensional analysis is to transform a problem into dimensionless form. The reason for this is that the approximation methods that are used to reduce difficult problems are based on comparisons. For example, in the projectile problem we simplified the differential equation by assuming that  $x$  was small compared to  $R$ . In contrast there are problems where the variable of interest is large, or it is slow or that it is fast compared to some other term in the problem. Whatever the comparison, it is important to know how all of the terms in the problem compare and for this we need the concept of scaling.

### 1.5.1 Projectile Problem

The reduction of the projectile equation (1.1) was based on the assumption that  $x$  is not very large, and so  $x + R$  could be replaced with just  $R$ . We will routinely use arguments like this to find an approximate solution and it is therefore essential we take more care in making such reductions. The way this is done is by first scaling the variables in the problem using characteristic values. The best way to explain what this means is to work out an example and the projectile problem is an excellent place to start.

#### 1.5.1.1 Change Variables

The first step in nondimensionalizing a problem is to introduce a change of variables, which for the projectile problem will have the form

$$t = t_c \tau,$$

$$x = x_c u.$$

In the above formula,  $x_c$  is a constant and it is a characteristic value of the variable  $x$ . It is going to be determined using the physical parameters in the problem, which for the projectile problem are  $g$ ,  $R$ , and  $v_0$ . In a similar manner,  $t_c$  is a constant that has the dimensions of time and it represents a characteristic value of the variable  $t$ . In some problems it will be clear at the beginning how to select  $x_c$  and  $t_c$ . However, it is assumed here that we have no clue at the start what to choose and will not select them until the problem is studied a bit more. All we know at the moment is that whatever the choice, the new variables  $u$ ,  $s$  are dimensionless. To make the change of variables note that from the chain rule

$$\begin{aligned}\frac{d}{dt} &= \frac{d\tau}{dt} \frac{d}{d\tau} \\ &= \frac{1}{t_c} \frac{d}{d\tau},\end{aligned}\tag{1.64}$$

and

$$\frac{d^2}{dt^2} = \frac{d}{dt} \left( \frac{d}{dt} \right) = \frac{1}{t_c^2} \frac{d^2}{d\tau^2}.\tag{1.65}$$

With this, the projectile equation (1.1) takes the form

$$\frac{1}{t_c^2} \frac{d^2}{d\tau^2} (x_c u) = - \frac{g R^2}{(R + x_c u)^2}.\tag{1.66}$$

The method requires us to collect the parameters into dimensionless groups. There is no unique way to do this, and this can cause confusion when first learning the procedure. For example, to nondimensionalize the denominator in (1.66) one can factor it as either  $R(1 + x_c u/R)$  or  $x_c(R/x_c + u)$ . The first has the benefit of enabling us to cancel the  $R$  in the numerator. Making this choice yields

$$\Pi_1 \frac{d^2 u}{d\tau^2} = - \frac{1}{(1 + \Pi_2 u)^2},\tag{1.67}$$

where the initial conditions (1.2) and (1.3) are

$$u(0) = 0,\tag{1.68}$$

$$\frac{du}{d\tau}(0) = \Pi_3.\tag{1.69}$$

In the above, the dimensionless groups are

$$\Pi_1 = \frac{x_c}{g t_c^2},\tag{1.70}$$

$$\Pi_2 = \frac{x_c}{R}, \quad (1.71)$$

$$\Pi_3 = \frac{t_c v_0}{x_c}. \quad (1.72)$$

### 1.5.1.2 The Dimensionless Groups

Our change of variables has resulted in three dimensionless groups appearing in the transformed problem. There are a few important points that need to be made here. First, the  $\Pi$ 's do not involve the variables  $u$ ,  $s$  and only depend on the parameters in the problem. Second, they are dimensionless and to accomplish this it was necessary to manipulate the projectile problem so the parameters end up grouped together to form dimensionless ratios. The third, and last, point is that the above three dimensionless groups are independent in the sense that it is not possible to write any one of them in terms of the other two. For example,  $\Pi_1$  is the only one that contains the parameter  $g$  while  $\Pi_2$  is the only one containing  $R$ . It is understood that in making the statement that the three groups are independent that  $x_c$  and  $t_c$  can be selected, if desired, independently of any of the parameters in the problem.

Before deciding on how to select  $x_c$  and  $t_c$ , it is informative to look a little closer at the above dimensionless groups. We begin with  $\Pi_2$ . In physical terms it is a measure of a typical, or characteristic, height of the projectile compared to the radius of the Earth. In comparison,  $\Pi_3$  is a measure of a typical, or characteristic, velocity  $x_c/t_c$  compared to the velocity the projectile starts with. Finally, the parameter group  $\Pi_1$  measures a typical, or characteristic, acceleration  $x_c/t_c^2$  in comparison to the acceleration due to gravity in a uniform field. These observations can be helpful when deciding on how to nondimensionalize a problem as will be shown next.

### 1.5.1.3 Use Dimensionless Groups to Determine Scaling

It is now time to actually decide on what to take for  $x_c$  and  $t_c$ . There are whole papers written on what to consider as you select these parameters, but we will take a somewhat simpler path. For our problem we have two parameters to determine, and we will do this by setting two of the above dimensionless groups equal to one. What we need to do is decide on which two to pick, and we will utilize what might be called rules of thumb.

**Rule of Thumb 1** Pick the  $\Pi$ 's that appear in the initial and/or boundary conditions.

We only have initial conditions in our problem, and the only dimensionless group involved with them is  $\Pi_3$ . So we set  $\Pi_3 = 1$  and conclude

$$x_c = v_0 t_c. \quad (1.73)$$



**Rule of Thumb 2** Pick the  $\Pi$ 's that appear in the reduced problem.

To use this rule it is first necessary to explain what the reduced problem is. This comes from the earlier assumption that the object does not get very high in comparison to the radius of the Earth, in other words,  $\Pi_2$  is small. The reduced problem is the one obtained in the extreme limit of  $\Pi_2 \rightarrow 0$ . Taking this limit in (1.67)–(1.69), and using (1.73), the reduced problem is

$$\Pi_1 \frac{d^2 u}{d\tau^2} = -1,$$

where

$$u(0) = 0, \quad \text{and} \quad \frac{du}{d\tau}(0) = 1.$$

According to the stated rule of thumb, we set  $\Pi_1 = 1$ , and so

$$x_c = v_0^2/g. \quad (1.74)$$

This choice for  $x_c$  seems reasonable based on our earlier conclusion that the maximum height for the uniform field case is  $v_0^2/(2g)$ .

Combining (1.73) and (1.74), we have that  $x_c = v_0^2/g$  and  $t_c = v_0/g$ . With this scaling, then (1.67)–(1.69) take the form

$$\frac{d^2 u}{d\tau^2} = -\frac{1}{(1 + \varepsilon u)^2}, \quad (1.75)$$

where

$$u(0) = 0, \quad (1.76)$$

$$\frac{du}{d\tau}(0) = 1. \quad (1.77)$$

The dimensionless parameter appearing in the above equation is

$$\varepsilon = \frac{v_0^2}{gR}. \quad (1.78)$$

This parameter will play a critical role in our constructing an accurate approximation of the solution of the projectile problem. This will be done in the next chapter but for the moment recall that since  $R \approx 6.4 \times 10^6$  m and  $g \approx 9.8$  m/s<sup>2</sup>, then  $\varepsilon \approx 1.6 \times 10^{-8} v_0^2$ . Consequently for baseball bats, sling shots, BB-guns, and other everyday projectile-producing situations, where  $v_0$  is not particularly large, the parameter  $\varepsilon$  is very small. This observation is central to the subject of the next chapter.

### 1.5.1.4 Changing Your Mind

Before leaving this example it is worth commenting on the nondimensionalization procedure by asking a question. Namely, how bad is it if different choices would have been made for  $x_c$  and  $t_c$ ? For example, suppose for some reason one decides to take  $\Pi_2 = 1$  and  $\Pi_3 = 1$ . The resulting projectile problem is

$$\varepsilon \frac{d^2 u}{d\tau^2} = -\frac{1}{(1+u)^2}, \quad (1.79)$$

where  $u(0) = 0$ ,  $\frac{du}{d\tau}(0) = 1$ , and  $\varepsilon$  is given in (1.78). No approximation has been made here and therefore this problem is mathematically equivalent to the one given in (1.75)–(1.77). Based on this, the answer to the question would be that using this other scaling is not so bad. However, the issue is amenability and what properties of the solution one is interested in. To explain, earlier we considered how the solution behaves if  $v_0$  is not very large. With the scaling that produced (1.79), small  $v_0$  translates into looking at what happens when  $\varepsilon$  is near zero. Unfortunately, the limit of  $\varepsilon \rightarrow 0$  results in the loss of the highest derivative in the differential equation and (1.79) reduces to  $0 = -1$ . How to handle such singular limits will be addressed in the next chapter but it requires more work than is necessary for this problem. In comparison, letting  $\varepsilon$  approach zero in (1.75) causes no such complications and for this reason it is more amenable to the study of the small  $v_0$  limit. The point here is that if there are particular limits, or conditions, on the parameters that it is worth taking them into account when constructing the scaling.

## 1.5.2 Weakly Nonlinear Diffusion

To explore possible extensions of the nondimensionalization procedure we consider a well-studied problem involving nonlinear diffusion. The problem consists of finding the concentration  $c(x, t)$  of a chemical for  $0 < x < \ell$ . The concentration satisfies

$$D \frac{\partial^2 c}{\partial x^2} = \frac{\partial c}{\partial t} - \lambda(\gamma - c)c, \quad (1.80)$$

where the boundary conditions are

$$c|_{x=0} = c|_{x=\ell} = 0, \quad (1.81)$$

and the initial condition is

$$c|_{t=0} = c_0 \sin(5\pi x/\ell). \quad (1.82)$$

The nonlinear diffusion equation (1.80) is known as Fisher's equation, and it arises in the study of the movement of genetic traits in a population. A common simplifying assumption made when studying this equation is that the nonlinearity is weak, which means that the term  $\lambda c^2$  is small in comparison to the others in the differential equation. This assumption will be accounted for in the nondimensionalization.

Before starting the nondimensionalization process we should look at the fundamental dimensions of the variables and parameters in the problem. First,  $c$  is a concentration, which corresponds to the number of molecules per unit volume, and so  $\llbracket c \rrbracket = L^{-3}$ . The units for the diffusion coefficient  $D$  were determined earlier, and it was found that  $\llbracket D \rrbracket = L^2/T$ . As for  $\gamma$ , the  $\gamma - c$  term in the differential equation requires these two quantities to have the same dimensions, and so  $\llbracket \gamma \rrbracket = \llbracket c \rrbracket$ . Similarly, from the differential equation we have  $\llbracket \lambda(\gamma - c)c \rrbracket = \llbracket \frac{\partial c}{\partial t} \rrbracket$ , and from this it follows that  $\llbracket \lambda \rrbracket = L^3 T^{-1}$ . Finally, from the initial condition we have that  $\llbracket c_0 \rrbracket = \llbracket c \rrbracket$ .

Now, to nondimensionalize the problem we introduce the change of variables

$$x = x_c y, \quad (1.83)$$

$$t = t_c \tau, \quad (1.84)$$

$$c = c_c u. \quad (1.85)$$

In this context,  $x_c$  has the dimensions of length and is a characteristic value of the variable  $x$ . Similar statements apply to  $t_c$  and  $c_c$ . Using the chain rule as in (1.64) the above differential equation takes the form

$$\frac{D c_c}{x_c^2} \frac{\partial^2 u}{\partial y^2} = \frac{c_c}{t_c} \frac{\partial u}{\partial \tau} - \lambda c_c (\gamma - c_c u) u.$$

It is necessary to collect the parameters into dimensionless groups, and so in the above equation we rearrange things a bit to obtain

$$\frac{D t_c}{x_c^2} \frac{\partial^2 u}{\partial y^2} = \frac{\partial u}{\partial \tau} - \lambda t_c c_c (\gamma / c_c - u) u. \quad (1.86)$$

In conjunction with this we have the boundary conditions

$$u|_{y=0} = u|_{y=\ell/x_c} = 0, \quad (1.87)$$

and the initial condition is

$$u|_{\tau=0} = (c_0/c_c) \sin(5\pi x_c y/\ell). \quad (1.88)$$

The resulting dimensionless groups are

$$\Pi_1 = \frac{Dt_c}{x_c^2}, \quad (1.89)$$

$$\Pi_2 = \lambda t_c c_c, \quad (1.90)$$

$$\Pi_3 = \gamma/c_c, \quad (1.91)$$

$$\Pi_4 = \ell/x_c, \quad (1.92)$$

$$\Pi_5 = c_0/c_c. \quad (1.93)$$

It is important to note that the five dimensionless groups given above are independent in the sense that it is not possible to write one of them in terms of the other four. As before, this statement is based on our ability to select, if desired, the scaling parameters  $x_c$ ,  $t_c$ ,  $c_c$  independently of each other and the other parameters in the problem. Also, in counting the dimensionless groups one might consider adding a sixth. Namely, in the initial condition (1.88) there is  $\Pi_6 = x_c/\ell$ . The reason it is not listed above is that it is not independent of the others because  $\Pi_6 = 1/\Pi_4$ .

We have three scaling parameters to specify, namely  $x_c$ ,  $t_c$ , and  $c_c$ . Using Rule of Thumb 1, the  $\Pi$ 's that appear in the boundary and initial conditions are set equal to one. In other words, we set  $\Pi_4 = 1$  and  $\Pi_5 = 1$ , from which it follows that  $x_c = \ell$  and  $c_c = c_0$ .

To use Rule of Thumb 2, we need to investigate what it means to say that the nonlinearity is weak. Equation (1.86) is nonlinear due to the term  $\lambda t_c c_c u^2 = \Pi_2 u^2$ , and the coefficient  $\Pi_2$  is the associated strength of the nonlinearity. For a weakly nonlinear problem one is interested in the solution for small values of  $\Pi_2$ . Taking the extreme limit we set  $\Pi_2 = 0$  in (1.86) to produce the reduced equation. The only group that remains in this limit is  $\Pi_1$ , and for this reason this is the group we select. So, setting  $\Pi_1 = 1$ , then we conclude  $t_c = \ell^2/D$ .

The resulting nondimensional diffusion equation is

$$\frac{\partial^2 u}{\partial y^2} = \frac{\partial u}{\partial \tau} - \varepsilon(b - u)u, \quad (1.94)$$

with boundary conditions

$$u(0, \tau) = u(1, \tau) = 0, \quad (1.95)$$

and the initial condition

$$u(y, 0) = \sin(5\pi y). \quad (1.96)$$

The dimensionless parameters appearing in the above equation are  $\varepsilon = \lambda c_0 \ell^2/D$  and  $b = \gamma/c_0$ . With this, weak nonlinearity corresponds to assuming that  $\varepsilon$  is small.

### 1.5.3 Endnotes

As you might have noticed, the assumption of a weak nonlinearity was used in the projectile problem. There are certainly other types of other extreme behavior. As an example, in nonlinear diffusion problems you come across situations involving weak diffusion. What this means for (1.86) is that  $Dt_c/x_c^2$  has a small value. In the extreme limit, which means this term is set to zero, then the only group that remains in the reduced problem is  $\Pi_2$ . Setting  $\Pi_2 = 1$ , then  $t_c = 1/(c_0\lambda)$ . With this, (1.86) becomes

$$\varepsilon \frac{\partial^2 u}{\partial y^2} = \frac{\partial u}{\partial \tau} - (b - u)u, \quad (1.97)$$

where  $\varepsilon = D/(c_0\lambda\ell^2)$  and  $b = \gamma/c_0$ . With this, weak diffusion corresponds to assuming that  $\varepsilon$  is small.

For those keeping track of the rules of thumb used to nondimensionalize a problem we have two. The first we ran across is the rule that the dimensionless groups in the initial and boundary conditions are set to one. The second rule arose when setting the dimensionless groups in the reduced problem to one. Although these can be effective rules, it is certainly possible to find problems where another scaling should be considered, and examples are given in Exercises 1.24 and 3.21. The overall objective in all cases is that the nondimensionalization is based on characteristic values of the variables.

## Exercises

### Section 1.2

**1.1** The example in Sect. 1.2.1 for the maximum height of a projectile was based on the assumption of a uniform gravitational field. In this exercise the nonuniform case is considered.

- (a) The projectile problem consists of solving (1.1), along with the initial conditions (1.2) and (1.3). There are three parameters in this problem, what are they? Assuming the maximum height  $x_M$  depends on these three parameters, show that the dimensionally reduced dependence has the form

$$x_M = \frac{v_0^2}{g} F(\Pi),$$

where  $\Pi$  is a dimensionless product.

- (b) It's possible to find the exact value of  $x_M$ . To do this, multiple (1.1) by  $x'(t)$  and then integrate both sides. Assuming the projectile does reach a maximum height, show that

$$x_M = \frac{v_0^2}{2g} \frac{1}{1 - Z},$$

where  $Z = v_0^2/(2Rg)$ . Use this to determine the function  $F$  in part (a).

**1.2** For a pendulum that starts from rest, the period  $p$  depends on the length  $\ell$  of the rod, on gravity  $g$ , on the mass  $m$  of the ball, and on the initial angle  $\theta_0$  at which the pendulum is started.

- Use dimensional analysis to determine the functional dependence of  $p$  on these four quantities.
- For the largest pendulum ever built, the rod is 70 ft and the ball weighs 950 lbs. Assuming that  $\theta_0 = \pi/6$  explain how to use a pendulum that fits on your desk to determine the period of this largest pendulum.
- Suppose it is found that  $p$  depends linearly on  $\theta_0$ , with  $p = 0$  if  $\theta_0 = 0$ . What does your result in part (a) reduce to in this case?

**1.3** The velocity  $v$  at which flow in a pipe will switch from laminar to turbulent depends on the diameter  $d$  of the pipe as well as on the density  $\rho$  and dynamic viscosity  $\mu$  of the fluid.

- Find a dimensionally reduced form for  $v$ .
- Suppose the pipe has diameter  $d = 100$  and for water (where  $\rho = 1$  and  $\mu = 10^{-2}$ ) it is found that  $v = 0.25$ . What is  $v$  for olive oil (where  $\rho = 1$  and  $\mu = 1$ )? The units here are in cgs.

**1.4** The luminosity of certain giant and supergiant stars varies in a periodic manner. It is hypothesized that the period  $p$  depends upon the star's average radius  $r$ , its mass  $m$ , and the gravitational constant  $G$ .

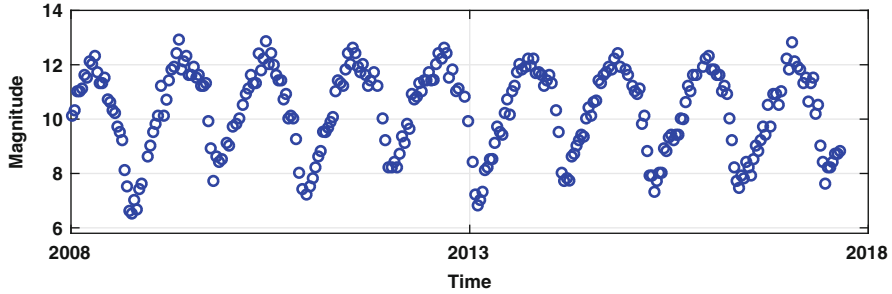
- Newton's law of gravitation asserts that the attractive force between two bodies is proportional to the product of their masses divided by the square of the distance between them, that is,

$$F = \frac{Gm_1m_2}{d^2},$$

where  $G$  is the gravitational constant. From this determine the (fundamental) dimensions of  $G$ .

- Use dimensional analysis to determine the functional dependence of  $p$  on  $m$ ,  $r$ , and  $G$ .
- Arthur Eddington used the theory for thermodynamic heat engines to show that

$$p = \sqrt{\frac{3\pi}{2\gamma G\rho}},$$



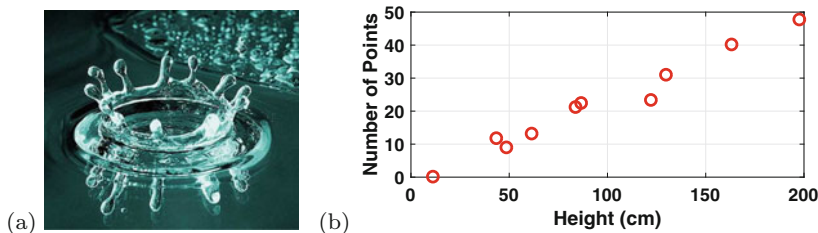
**Fig. 1.8** Luminosity of a Mira type variable star, 1621+19 U Herculis (AAVSO (2018))

where  $\rho$  is the average density of the star and  $\gamma$  is the ratio of specific heats for stellar material. What must be assumed so your result from part (b) yields this formula?

- (d) In Fig. 1.8 the data for a pulsating star are given. Explain how you could use data like this to complete the formula you derived in part (b).

**1.5** When a drop of liquid hits a wetted surface a crown formation appears, as shown in Fig. 1.9a. It has been found that the number of points  $N$  on the crown depends on the speed  $U$  at which the drop hits the surface, the radius  $r$  and density  $\rho$  of the drop, and the surface tension  $\sigma$  of the liquid making up the drop. How  $N$  depends on these quantities has been studied extensively and some of the reasons why are given in Rioboo et al. (2003).

- Use dimensional reduction to determine the functional dependence of  $N$  on  $U$ ,  $r$ ,  $\rho$ , and  $\sigma$ . Express your answer in terms of the Weber number  $W_e = \rho U^2 r / \sigma$ .
- The value of  $N$  has been measured as a function of the initial height  $h$  of the drop and the results are shown in Fig. 1.9b. Express your answer in part (a) in terms of  $h$  by writing  $U$  in terms of  $h$  and  $g$ . Assume the drop starts with zero velocity.
- The data in Fig. 1.9b show that  $N$  is zero for small  $h$ , but once the height is large enough, then  $N$  grows linearly with  $h$ . Use this, and your result from part (b), to find the unknown function in part (a). In the experiments,  $r = 3.6$  mm,  $\rho = 1.1014$  gm/cm<sup>3</sup>, and  $\sigma = 50.5$  dyn/cm.
- According to your result from part (c), what must the initial height of the drop be to produce at least 80 points?
- According to your result from part (c), how many points are generated for a drop of mercury when  $h = 200$  cm? Assume  $r = 3.6$  mm,  $\rho = 13.5$  gm/cm<sup>3</sup>, and  $\sigma = 435$  dyn/cm.



**Fig. 1.9** (a) Formation of a crown when a liquid drop hits a wetted surface. (b) The measured values of the number of points  $N$  (Hobbs and Kezweent 1967)

**1.6** The frequency  $\omega$  of waves on a deep ocean is found to depend on the wavelength  $\lambda$  of the wave, the surface tension  $\sigma$  of the water, the density  $\rho$  of the water, and gravity.

- Use dimensional reduction to determine the functional dependence of  $\omega$  on  $\lambda$ ,  $\sigma$ ,  $\rho$ , and  $g$ .
- In fluid dynamics it is shown that (see Exercise 9.24)

$$\omega = \sqrt{gk + \frac{\sigma k^3}{\rho}},$$

where  $k = 2\pi/\lambda$  is the wavenumber. Show how to obtain this using your result from part (a).

**1.7** A ball is dropped from a height  $h_0$  and it rebounds to a height  $h_r$ . The rebound height depends on the elastic modulus  $E$ , radius  $R$ , and the mass density  $\rho$  of the ball. It also depends on the initial height  $h_0$  as well as on the gravitational acceleration constant  $g$ .

- Find a dimensionally reduced form for  $h_r$ .
- Suppose it is found that  $h_r$  depends linearly on  $h_0$ , with  $h_r = 0$  if  $h_0 = 0$ . What does your formula from part (a) reduce to in this case?
- Suppose the density of the ball is doubled. Use the result in (a) to explain how to change  $E$  so the rebound height stays the same.

**1.8** A ball, when released underwater, will rise towards the surface with velocity  $v$ . This velocity depends on the density  $\rho_b$  and radius  $R$  of the ball, on gravity  $g$ , and on the density  $\rho_f$  and kinematic viscosity  $\nu$  of the water.

- Find a dimensionally reduced form for  $v$ .
- In fluid mechanics, using Stokes' Law, it is found that

$$v = \frac{2gR^2(\rho_b - \rho_f)}{9\nu\rho_f}.$$



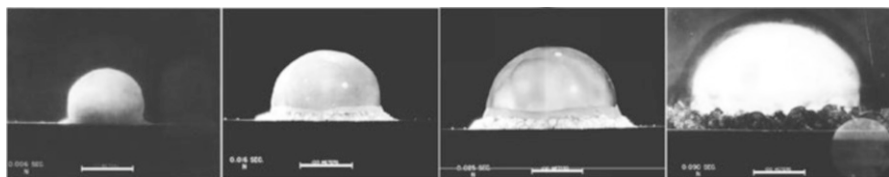
How does this differ from your result from part (a)? It is interesting to note that this formula is used by experimentalists to determine the viscosity of fluids. They do this by measuring the velocity in an apparatus called a falling ball viscometer, and then use the above formula to determine  $\nu$ .

**1.9** The velocity  $v$  of water through a circular pipe depends on the pressure difference  $p$  between the two ends of the pipe, the length  $\ell$  and radius  $r$  of the pipe, as well as on the dynamic viscosity  $\mu$  and density  $\rho$  of the water.

- Use dimensional analysis to determine the functional dependence of  $v$  on the above quantities.
- Suppose it is found that  $v$  depends linearly on  $p$ , with  $v = 0$  if  $p = 0$ . What does your formula from part (a) reduce to in this case?
- Your formula from part (b) should contain a general function of one, or more, dimensionless products. Explain how to experimentally determine this function. Be specific about which parameters are fixed, and which are varied, in the experiment. Also, your experiment should vary as few of the parameters as possible in determining this function.

**1.10** In a high energy explosion there is a very rapid release of energy  $E$  that produces an approximately spherical shock wave that expands in time (Fig. 1.10).

- Assuming the radius  $R$  of the shock wave depends on  $E$ , the length of time  $t$  since the explosion, and the density  $\rho$  of the air, use dimensional reduction to determine how the radius depends on these quantities. This expression is known as the Taylor-Sedov formula.
- It was shown by G. I. Taylor that if  $E = 1\text{ J}$  and  $\rho = 1\text{ kg/m}^3$ , then  $R = t^{2/5}\text{ m/s}^{2/5}$ . Use this information and the result from (a) to find the exact formula for  $R$ .
- Use the photographs in Fig. 1.10, and your result from (b) to estimate the energy released. The air density is  $\rho = 1\text{ kg/m}^3$ .
- The blast wave from a supernova can be modeled using the Taylor-Sedov formula. Explain how this can be used to estimate the date the supernova took place, using your result from part (b). As an example, use Tycho, which currently has a radius of about 33.2 light years, an estimated energy of  $10^{44}\text{ J}$ , and density  $\rho = 2 \times 10^{-21}\text{ kg/m}^3$ .



**Fig. 1.10** Shock wave produced by a nuclear explosion, at 6 ms, 16 ms, 25 ms, and 90 ms. The width of the white bar in each figure is 100 m (Brixner (2018))

**1.11** In quantum chromodynamics three parameters that play a central role are the speed of light  $c$ , Planck's constant  $\hbar$ , and the gravitational constant  $G$ .

- Explain why it is possible to use  $\llbracket c \rrbracket$ ,  $\llbracket \hbar \rrbracket$ ,  $\llbracket G \rrbracket$  as fundamental units.
- The distance  $\ell_p$  at which the strong, electromagnetic, and weak forces become equal depends on  $c$ ,  $\hbar$ ,  $G$ . Find a dimensionally reduced form for how  $\ell_p$  depends on these three parameters. Based on this result, if the speed of light were to double what happens to  $\ell_p$ ?
- The Bohr radius  $a$  of an electron depends on  $\hbar$ , the electron's charge  $e$ , and the mass  $m_e$  of the electron. Find a dimensionally reduced form for  $a$ .

**1.12** The speed  $c_m$  at which magnetosonic waves travel through a plasma depends on the magnitude  $B$  of the magnetic field, the permeability  $\mu_0$  of free space, and the density  $\rho$  and pressure  $p$  of the plasma.

- Use dimensional reduction to determine the functional dependence of  $c_m$  on  $B$ ,  $\mu_0$ ,  $\rho$ , and  $p$ .
- From the basic laws for plasmas it is shown that

$$c_m = \sqrt{V_A^2 + c_s^2},$$

where  $V_A = B\sqrt{\mu_0/\rho}$  is the Alfvén speed and  $c_s = \sqrt{\gamma p/\rho}$  is the sound speed in the gas. In the latter expression,  $\gamma$  is a number. How does this differ from your result in (a)?

## Section 1.4

**1.13** Suppose the mass density  $u(x, t)$  satisfies the partial differential equation

$$u_t + Bu_{xxxx} = 0, \quad \text{for } \begin{cases} 0 < x < \infty, \\ 0 < t, \end{cases}$$

where  $u = u_0$  and  $u_x = 0$  at  $x = 0$ , while  $u \rightarrow 0$  and  $u_x \rightarrow 0$  as  $x \rightarrow \infty$ . Also,  $u = 0$  at  $t = 0$ . Assume that the constant  $B$  is positive, and  $u_0$  is nonzero.

- Use dimensional reduction, and a similarity variable, to reduce the partial differential equation to an ordinary differential equation.
- Write the original boundary and initial conditions in terms of the similarity variable.

**1.14** Suppose the temperature  $h(x, t)$  satisfies the partial differential equation

$$\frac{\partial h}{\partial t} = \kappa \frac{\partial}{\partial x} \left( h^3 \frac{\partial h}{\partial x} \right), \quad \text{for } \begin{cases} 0 < x < \infty, \\ 0 < t, \end{cases}$$

where  $h = h_0$  at  $x = 0$ ,  $h \rightarrow 0$  as  $x \rightarrow \infty$ , and  $h = 0$  at  $t = 0$ . Assume that the constants  $\kappa$  and  $h_0$  are positive.

- What are the dimensions of  $\kappa$ ?
- Use dimensional reduction, and a similarity variable, to reduce the partial differential equation to an ordinary differential equation.
- Write the original boundary and initial conditions in terms of the similarity variable.

**1.15** In the study of the motion of particles moving along the  $x$ -axis one comes across the problem of finding the velocity  $u(x, t)$  that satisfies the nonlinear partial differential equation

$$u_t + uu_x = 0, \quad \text{for } \begin{cases} -\infty < x < \infty, \\ 0 < t, \end{cases} \quad (1.98)$$

where

$$u(x, 0) = \begin{cases} 0 & \text{if } x < 0 \\ u_0 & \text{if } 0 < x. \end{cases} \quad (1.99)$$

Assume that  $u_0$  is a positive constant. Equation (1.98) is derived in Chap. 5, and it is known as the inviscid Burgers' equation. It, along with the jump condition in (1.99), form what is known as a Riemann problem.

- What three physical quantities does  $u$  depend on?
- Use dimensional reduction, and a similarity variable, to reduce this problem to a nonlinear ordinary differential equation with two boundary conditions.
- Use the result from part (b) to solve the Riemann problem. The solution, which is known as an expansion fan, must be continuous for  $t > 0$ .
- What is the solution if the initial condition (1.99) is replaced with  $u(x, 0) = u_0$ ?
- Suppose that, rather than velocity, the variable  $u$  is displacement. Explain why it is not possible for  $u$  to satisfy (1.98).

**1.16** Suppose the temperature  $u(x, t)$  satisfies the diffusion equation

$$D \frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t}, \quad \text{for } \begin{cases} -\infty < x < \infty, \\ 0 < t, \end{cases}$$

where the boundary conditions are

$$u = 0, \quad \text{as } x \rightarrow \pm\infty.$$

Instead of an initial condition, assume the solution satisfies

$$\int_{-\infty}^{\infty} u dx = \gamma, \quad \forall t > 0. \quad (1.100)$$

- (a) What are the dimensions of  $\gamma$ ?
- (b) What four physical quantities does  $u(x, t)$  depend on? Use this to find a dimensionally reduced form for  $u(x, t)$ .
- (c) Using the result from part (b), transform the diffusion equation into an ordinary differential equation. How do the boundary conditions transform?
- (d) Find the solution of the problem from part (c). You can assume  $F' \rightarrow 0$  and  $\eta F \rightarrow 0$  as  $\eta \rightarrow \pm\infty$ . As a hint, you might want to look for the expression  $(\eta F)'$  in your equation.
- (e) Rewrite (1.100) in terms of  $F$ . Using this to find the arbitrary constant in your solution from part (d), show that

$$u = \frac{\gamma}{2\sqrt{\pi Dt}} e^{-x^2/(4Dt)}.$$

This is known as the fundamental, or point source, solution of the diffusion equation.

## Section 1.5

**1.17** The vertical displacement  $u(x)$  of an elastic string of length  $\ell$  satisfies the boundary value problem

$$\tau \frac{d^2 u}{dx^2} + \mu u = p, \quad \text{for } 0 < x < \ell,$$

where  $u(0) = 0$  and  $u(\ell) = U$ . Also,  $p$  is a constant and it has the dimensions of force per length.

- (a) What are the dimensions for the constants  $\tau$  and  $\mu$ ?
- (b) Show how it is possible to nondimensionalize this problem so it takes the form

$$\frac{d^2 v}{ds^2} + \alpha v = \beta, \quad \text{for } 0 < s < 1,$$

where  $v(0) = 0$  and  $v(1) = 1$ . Make sure to state what  $\alpha$  and  $\beta$  are.

**1.18** From Newton's second law, the displacement  $y(t)$  of the mass in a mass, spring, dashpot system satisfies

$$m \frac{d^2 y}{dt^2} = F_s + F_d,$$

where  $m$  is the mass,  $F_s$  is the restoring force in the spring, and  $F_d$  is the damping force. To have a complete IVP we need to state the initial conditions, and for this problem assume

$$y(0) = 0, \quad \frac{dy}{dt}(0) = v_0.$$

- (a) Suppose there is no damping, so  $F_d = 0$ , and the spring is linear, so  $F_s = -ky$ . What are the dimensions for the spring constant  $k$ ? Nondimensionalize the resulting IVP. Your choice for  $y_c$  and  $t_c$  should result in no dimensionless products being left in the IVP.
- (b) Now, in addition to a linear spring, suppose linear damping is included, so,

$$F_d = -c \frac{dy}{dt}.$$

What are the dimensions for the damping constant  $c$ ? Using the same scaling as in part (a), nondimensionalize the IVP. Your answer should contain a dimensionless parameter  $\varepsilon$  that measures the strength of the damping. In particular, if  $c$  is small, then  $\varepsilon$  is small. The system in this case is said to have weak damping.

**1.19** The velocity  $v(t)$  of a sphere dropping through the atmosphere satisfies

$$m \frac{dv}{dt} = -mg + D_F,$$

where  $m$  is the mass of the sphere,  $g$  is the gravitational acceleration constant, and  $D_F$  is the drag force. According to (1.19),  $D_F = \rho R^2 v^2 G(Re)$ , where  $Re = R|v|\rho/\mu$ . A reasonable approximation based on the data shown in Fig. 1.3, which is to be used in this exercise, is  $G(Re) = \frac{2}{3} + \frac{24}{Re}$ . It is assumed that  $v(0) = 0$ , and upward is the positive direction (so,  $v(t) \leq 0$ ).

- (a) Assuming a weak nonlinearity, use the Rules of Thumb given in Sect. 1.5 to show that a nondimensional version of the problem is

$$\frac{du}{d\tau} = -1 - u + \varepsilon u^2,$$

where  $u(0) = 0$ .

- (b) What about the sphere must hold for the nonlinearity to be weak?

**1.20** The Newton-Sefan law of cooling states that the temperature  $T(t)$  is determined by solving

$$\frac{dT}{dt} = -k_0(T - T_a) - k_1(T^4 - T_a^4),$$

for  $T(0) = T_0$ . The constant  $T_a$  is the ambient temperature and, since the object is cooling,  $T_0 > T_a$ .

- Letting  $T(t) = T_a + h(t)$ , find the differential equation and initial condition that  $h(t)$  satisfies.
- Show how it is possible to nondimensionalize the problem for  $h(t)$  so it has the form

$$\frac{du}{d\tau} = -u - \frac{\alpha}{\varepsilon}[(1 + \varepsilon u)^4 - 1],$$

where  $u(0) = 1$  and  $\alpha$  is a positive constant.

- What assumption is being made so that  $\varepsilon$  is small?

**1.21** An equation for the displacement  $u(x, t)$  of a string, on the interval  $0 < x < \ell$ , is (Chen and Ding, 2008)

$$c^2 \frac{\partial^2 u}{\partial t^2} = \left[ 1 + \left( \frac{\partial u}{\partial x} \right)^2 \right] \frac{\partial^2 u}{\partial x^2},$$

where  $u(0, t) = u_0$ ,  $u(\ell, t) = 0$ ,  $u(x, 0) = 0$ , and  $u_t(x, 0) = 0$ . Assume that  $c$  is a positive constant.

- What are the dimensions of  $c$  and  $u_0$ ?
- Assuming a weak nonlinearity, use the Rules of Thumb given in Sect. 1.5 to nondimensionalize this problem.

**1.22** The equation for the displacement  $u(x, t)$  of an elastic beam, on the interval  $0 < x < \ell$ , is

$$EI \frac{\partial^4 u}{\partial x^4} + \rho \frac{\partial^2 u}{\partial t^2} = 0,$$

where the boundary conditions are  $u = u_0 \sin(\omega t)$  and  $\frac{\partial u}{\partial x} = 0$  at  $x = 0$ , while  $u = \frac{\partial u}{\partial x} = 0$  at  $x = \ell$ . Assume the initial conditions are  $u = 0$  and  $\frac{\partial u}{\partial t} = 0$  at  $t = 0$ . Here  $E$  is the elastic modulus,  $I$  is the moment of inertia, and  $\rho$  is the mass per unit length of the beam. Nondimensionalize the problem in such a way that the resulting boundary conditions contain no nondimensional groups.

**1.23** The equation for the velocity  $v(x, t)$  of a fluid, on the interval  $0 < x < \ell$ , is

$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = \nu \frac{\partial^2 v}{\partial x^2},$$

where  $v(0, t) = 0$ ,  $v(\ell, t) = 0$ , and  $v(x, 0) = v_0 x(\ell - x)/\ell^2$ . Assume that  $\nu$  and  $v_0$  are positive constants.

- What are the dimensions of  $\nu$  and  $v_0$ ?

- (b) Assuming a weak nonlinearity, use the Rules of Thumb given in Sect. 1.5 to nondimensionalize this problem.

**1.24** A thermokinetic model for the concentration  $u$  and temperature  $q$  of a mixture consists of the following equations (Gray and Scott 1994):

$$\begin{aligned}\frac{du}{dt} &= k_1 - k_2 u e^{k_3 q}, \\ \frac{dq}{dt} &= k_4 u e^{k_3 q} - k_5 q.\end{aligned}$$

The initial conditions are  $u = 0$  and  $q = 0$  at  $t = 0$ .

- What are the dimensions of each of the  $k_i$ 's?
- Explain why the rule of thumb for scaling used in the projectile problem does not help here.
- Find the steady-state solution, that is, the solution of the differential equations with  $u' = 0$  and  $q' = 0$ .
- Nondimensionalize the problem using the steady-state solution from (c) to scale  $u$  and  $q$ . Make sure to explain how you selected the scaling for  $t$ .

**1.25** The equations that account for the relativistic effects on the orbit of a planet around the sun are

$$\begin{aligned}\frac{d}{dt}\left(\gamma \frac{dr}{dt}\right) &= -\frac{Gm}{r^2} + \gamma r \left(\frac{d\theta}{dt}\right)^2, \\ \frac{d}{dt}\left(\gamma r^2 \frac{d\theta}{dt}\right) &= 0,\end{aligned}$$

where  $\gamma = \sqrt{1 - v^2/c^2}$  and  $v^2 = r_t^2 + r^2 \theta_t^2$ . Also,  $c$  is the speed of light,  $G$  is the gravitational constant, and  $m$  is the mass of the sun. Assume the initial conditions are  $r(0) = r_0$ ,  $r'(0) = 0$ ,  $\theta(0) = 0$ , and  $\theta'(0) = \dot{\theta}_0$ . The nonrelativistic equations are obtained in the limit of  $v/c \rightarrow 0$ .

- What are the dimensions of  $r_0$  and  $\dot{\theta}_0$ ?
- Nondimensionalize the problem based on the assumption of weak relativistic effects.

## Additional Questions

**1.26** This problem considers various questions that can arise in dimensional analysis.

- (a) In the drag on a sphere example, suppose that it is known that  $D_F$  is a linear function of  $R$ . How does this simplify the general solution?
- (b) In the drag on a sphere example, suppose that using SI units  $\mu = 5$ ,  $R = 1$ ,  $v = 1$ , and  $\rho = 1$ . In this case,  $\Pi = 5$ , where the formula for  $\Pi$  is given in (1.15). What is the value of  $\Pi$  using CGS (centimeter-gram-second system of units)? What is the value of  $\Pi$  using USC (United States customary units)?
- (c) In the toppling dominoes example, once the general product solution (1.26) is derived, the question arises as what the general solution might be. A very common guess is that  $v = \sqrt{hg} F_1(\Pi_1) F_2(\Pi_2)$ , where  $F_1$  and  $F_2$  are arbitrary functions. Show that this cannot be the general solution by giving an example for  $v$  that cannot be written in this way.

**1.27** This problem considers various properties of the  $\llbracket \cdot \rrbracket$  operator.

- (a) Given physical quantities  $x$  and  $y$ , explain why the following hold: i)  $\llbracket xy \rrbracket = \llbracket x \rrbracket \llbracket y \rrbracket$ , ii) if  $\llbracket y \rrbracket = \llbracket x \rrbracket$ , then  $\llbracket x + y \rrbracket = \llbracket x \rrbracket$ , and iii) for any number  $\alpha$ ,  $\llbracket \alpha x \rrbracket = \llbracket x \rrbracket$ . You can assume here, if you wish, that  $M$ ,  $L$ , and  $T$  are the only fundamental dimensions involved.
- (b) Explain why the formulas in (1.8) and (1.9) can be written in the simple statement that

$$\llbracket \frac{d}{dt} \rrbracket = \frac{1}{\llbracket t \rrbracket}.$$

- (c) Using the results from parts (a) and (b), one obtains

$$\llbracket \partial_x(u \partial_t u) \rrbracket = \llbracket \partial_x \rrbracket \llbracket u \rrbracket \llbracket \partial_t \rrbracket \llbracket u \rrbracket = \frac{\llbracket u \rrbracket^2}{\llbracket x \rrbracket \llbracket t \rrbracket}.$$

Use a similar approach to find  $\llbracket \partial_x^2(u^3 \partial_t u) \rrbracket$  and  $\llbracket \partial_x(\partial_t u^7) \rrbracket$ .

- (d) Using the definition of a definite integral, determine  $\llbracket \int_a^b f(t) dt \rrbracket$  in terms of  $\llbracket f \rrbracket$  and  $\llbracket t \rrbracket$ .

**1.28** This problem explores some consequences of dimensional quantities.

- (a) If  $g$  is the gravitational acceleration constant, explain why  $\sin(g)$  and  $e^g$  make no sense.
- (b) Explain why density, volume, and velocity can be used in place of length, mass, and time as fundamental units.
- (c) Explain why volume, velocity, and acceleration cannot be used in place of length, mass, and time as fundamental units.

**1.29** In some cases where the similarity method fails to simplify the problem, it is possible to use other properties of the solution, so the method works. This is demonstrated for the diffusion problem (1.47)–(1.49).



- (a) Suppose  $u$  has the dimensions of velocity. Show that dimensional reduction produces two similarity variables (hence, there is no reduction in the number of independent variables).
- (b) The key observation is that if you increase  $u_0$  by a factor  $\alpha$ , then the solution  $u$  increases by this same factor. To prove this, let  $U(x, t)$  be the solution for a given  $u_0$ . Setting  $u = \alpha U(x, t)$ , what problem does  $u$  satisfy? Assuming the solution is unique, explain why this proves the key observation.
- (c) Explain why the key observation made in part (b) means that  $d = 1$  in (1.52).
- (d) Suppose  $u$  has the dimensions of velocity. Use a similarity reduction to find the solution of the problem.

**1.30** The problem involves finding the velocity  $u(x, t)$  of the waves in shallow water. It is assumed to satisfy the linearized KdV equation, which is

$$\frac{\partial u}{\partial t} + K \frac{\partial^3 u}{\partial x^3} = 0, \quad \text{for } \begin{cases} -\infty < x < \infty, \\ 0 < t, \end{cases}$$

where  $K$  is a positive constant. The boundary conditions are  $u = 0$  as  $x \rightarrow \pm\infty$ , and instead of an initial condition, assume the solution satisfies

$$\int_{-\infty}^{\infty} u dx = \gamma, \quad \forall t > 0. \quad (1.101)$$

- (a) What are the dimensions of  $K$  and  $\gamma$ ?
- (b) What four physical quantities does  $u(x, t)$  depend on? With this, write down the equivalent version of (1.52) for this problem, and use the exponent  $d$  for  $\gamma$ .
- (c) The key observation is that if you increase  $\gamma$  by a factor  $\alpha$ , then the solution  $u$  increases by this same factor. To prove this, let  $U(x, t)$  be the solution for a given  $\gamma$ . Setting  $u = \alpha U(x, t)$ , what problem does  $u$  satisfy? Assuming the solution is unique, explain why this proves the key observation.
- (d) Explain why the key observation made in part (c) means that  $d = 1$ . With this, find the general dimensionally reduced form of the solution. As usual, let  $\eta$  denote the similarity variable.
- (e) Using the result from part (d), transform the KdV equation into an ordinary differential equation. How do the boundary conditions transform?
- (f) Show that the third order differential equation in part (e) can be simplified to one that is second order. You can assume  $F'' \rightarrow 0$  and  $\eta F \rightarrow 0$  as  $\eta \rightarrow \pm\infty$ . As a hint, you might want to look for the expression  $(\eta F)'$  in your equation.
- (g) Rewrite (1.101) in terms of  $F$ .

Comment: The solution of the problem is  $F(\eta) = \text{Ai}(\eta)$ , where  $\text{Ai}$  is known as the Airy function. This function plays a central role in wave problems, both in fluids, electromagnetics, and quantum mechanics (Vallée and Soares, 2010).

**1.31** When an end of a slender strip of paper is put into a cup of water, because of absorption, the water rises up the paper. The density  $\rho$  of the water along the strip satisfies the differential equation

$$\frac{\partial \rho}{\partial t} + \frac{\partial J}{\partial x} = 0,$$

where  $J$  is known as the flux.

- (a) What are the dimensions of  $J$ ?
- (b) The flux  $J$  depends on the gravitational constant  $g$ , the strip width  $d$ , the density gradient  $\frac{\partial \rho}{\partial x}$ , and the surface tension  $\sigma$  of the water. Find a dimensionally reduced form for  $J$ .
- (c) What does your result in (b) reduce to if it is found that  $J$  depends linearly on the density gradient, with  $J = 0$  if  $\rho_x = 0$ ? What is the resulting differential equation?
- (d) If the strip has length  $h$  the boundary conditions are  $\rho = \rho_0$  at  $x = 0$  and  $J = 0$  at  $x = h$ . The initial condition is  $\rho = 0$  at  $t = 0$ . With this, and your differential equation from (c), nondimensionalize the problem for  $\rho$  in such a way that no nondimensional groups appear in the final answer.

# Chapter 2

## Perturbation Methods



### 2.1 Regular Expansions

To introduce the ideas underlying perturbation methods and asymptotic approximations, we will begin with an algebraic equation. The problem we will consider is how to find an accurate approximation of the solution  $x$  of the quadratic equation

$$x^2 + 2\varepsilon x - 1 = 0, \quad (2.1)$$

in the case of when  $\varepsilon$  is a small positive number. The examples that follow this one are more complex and, unlike this equation, we will not necessarily know at the start how many solutions the equation has. A method for determining the number of real-valued solutions involves sketching the terms in the equation. With this in mind, we rewrite the equation as  $x^2 - 1 = -2\varepsilon x$ . The left- and right-hand sides of this equation are sketched in Fig. 2.1. Based on the intersection points, it is seen that there are two solutions. One is a bit smaller than  $x = 1$  and the other is just to the left of  $x = -1$ . Another observation is that the number of solutions does not change as  $\varepsilon \rightarrow 0$ . The fact that the reduced problem, which is the one obtained when setting  $\varepsilon = 0$ , has the same number of solutions as the original problem is a hallmark of what are called regular perturbation problems.

Our goal is to derive approximations of the solutions for small  $\varepsilon$ , and for this simple problem we have a couple of options on how to do this.

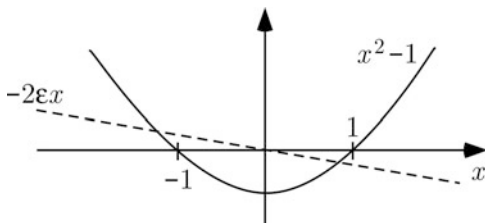
#### Method 1: Solve Then Expand

It is an easy matter to find the solution using the quadratic formula. The result is

$$x = -\varepsilon \pm \sqrt{1 + \varepsilon^2}. \quad (2.2)$$

This completes the solve phase of the process. To obtain an approximation for the two solutions, for small  $\varepsilon$ , we first use the binomial expansion (see Table 2.1) to obtain

**Fig. 2.1** Sketch of the functions appearing in the quadratic equation in (2.1)



$$\sqrt{1 + \varepsilon^2} = 1 + \frac{1}{2}\varepsilon^2 - \frac{1}{8}\varepsilon^4 + \dots \quad (2.3)$$

The series on the right converges for  $\varepsilon^2 < 1$ , which holds for our problem as we are assuming that  $\varepsilon$  is close to zero.

Substituting (2.3) into (2.2) yields

$$\begin{aligned} x &= -\varepsilon \pm \left( 1 + \frac{1}{2}\varepsilon^2 - \frac{1}{8}\varepsilon^4 + \dots \right) \\ &= \pm 1 - \varepsilon \pm \frac{1}{2}\varepsilon^2 \mp \frac{1}{8}\varepsilon^4 + \dots \end{aligned} \quad (2.4)$$

In the last step the terms are listed in order according to their power of  $\varepsilon$ . With this we can list various levels of approximation of the solutions, as follows

$x \approx \pm 1$	1 term approximation
$x \approx \pm 1 - \varepsilon$	2 term approximation
$x \approx \pm 1 - \varepsilon \pm \frac{1}{2}\varepsilon^2$	3 term approximation.

So, we have accomplished what we set out to do, which is to derive an approximation of the solution for small  $\varepsilon$ . The procedure is straightforward but it has a major drawback because it requires us to be able to first solve the equation before constructing the approximation. For most problems this is simply impossible, so we need another approach.

An important comment is needed about how (2.3) was derived. Although it is possible to obtain (2.3) using a straightforward version of a Taylor series, it is much easier to use the substitution property for a Taylor series. Those unfamiliar with this should review the examples in Sect. A.1.1. This is recommended as substitution is used on numerous occasions in this textbook, and this is often done without explicit mention that substitution is used.

### Method 2: Expand Then Solve

This approach requires us to first state what we consider to be the general form of the approximation for  $x$ . This requires a certain amount of experience and a reasonable place to start is with Taylor's theorem. We know that the solution depends on  $\varepsilon$ , we just don't know how. Emphasizing this dependence by writing  $x(\varepsilon)$ , then using

**Table 2.1** Taylor series expansions, about  $x = 0$ , for some of the more commonly used functions

---

$f(x) = f(0) + xf'(0) + \frac{1}{2}x^2f''(0) + \frac{1}{3!}x^3f'''(0) + \dots$
$(a+x)^\gamma = a^\gamma + \gamma xa^{\gamma-1} + \frac{1}{2}\gamma(\gamma-1)x^2a^{\gamma-2} + \frac{1}{3!}\gamma(\gamma-1)(\gamma-2)x^3a^{\gamma-3} + \dots$
$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \dots$
$\frac{1}{(1+x)^2} = 1 - 2x + 3x^2 - 4x^3 + \dots$
$\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 + \dots$
$\frac{1}{\sqrt{1+x}} = 1 - \frac{1}{2}x + \frac{3}{8}x^2 - \frac{5}{16}x^3 + \dots$
$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \dots$
$\sin(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots$
$\cos(x) = 1 - \frac{1}{2}x^2 + \frac{1}{4!}x^4 + \dots$
$\sin(a+x) = \sin(a) + x \cos(a) - \frac{1}{2}x^2 \sin(a) - \frac{1}{6}x^3 \cos(a) + \dots$
$\cos(a+x) = \cos(a) - x \sin(a) - \frac{1}{2}x^2 \cos(a) + \frac{1}{6}x^3 \sin(a) + \dots$
$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 + \dots$
$\ln(a+x) = \ln(a) + \ln(1+x/a) = \ln(a) + \frac{x}{a} - \frac{1}{2}\left(\frac{x}{a}\right)^2 + \frac{1}{3}\left(\frac{x}{a}\right)^3 + \dots$

---

Taylor's theorem for  $\varepsilon$  near zero, we obtain

$$x(\varepsilon) = x(0) + \varepsilon x'(0) + \frac{1}{2}\varepsilon^2 x''(0) + \dots \quad (2.5)$$

This implies that  $x$  can be expanded using integer powers of  $\varepsilon$ . Assuming this applies to the solution, then the expansion has the form

$$x \sim x_0 + x_1\varepsilon + x_2\varepsilon^2 + \dots \quad (2.6)$$

We will need to be able to identify the coefficients in an expansion and the big  $O$  notation will be used for this. So, in (2.6) the  $O(\varepsilon)$  coefficient is  $x_1$  and the  $O(\varepsilon^2)$  coefficient is  $x_2$ . For the same reason,  $x_0$  is the coefficient of the  $O(1)$  term.

We will substitute (2.6) into (2.1), but before doing so note

$$\begin{aligned} x^2 &\sim (x_0 + x_1\varepsilon + x_2\varepsilon^2 + \cdots)(x_0 + x_1\varepsilon + x_2\varepsilon^2 + \cdots) \\ &\sim x_0^2 + 2x_0x_1\varepsilon + (2x_0x_2 + x_1^2)\varepsilon^2 + \cdots. \end{aligned} \quad (2.7)$$

With this, (2.1) takes the form

$$x_0^2 + 2x_0x_1\varepsilon + (2x_0x_2 + x_1^2)\varepsilon^2 + \cdots + 2\varepsilon(x_0 + x_1\varepsilon + \cdots) - 1 = 0,$$

which can be rearranged in the form

$$x_0^2 - 1 + 2(x_0x_1 + x_0)\varepsilon + (2x_0x_2 + x_1^2 + 2x_1)\varepsilon^2 + \cdots = 0. \quad (2.8)$$

We are constructing an approximation for small  $\varepsilon$ . So, in letting  $\varepsilon \rightarrow 0$  in the above equation we obtain the equation for the  $O(1)$  term.

$$O(1) \quad x_0^2 - 1 = 0$$

The solutions are  $x_0 = \pm 1$ .

There are a couple of ways to now determine  $x_1$ . The mathematically more complete approach is to notice that (2.8) reduces to

$$2(x_0x_1 + x_0) + (2x_0x_2 + x_1^2 + 2x_1)\varepsilon + \cdots = 0. \quad (2.9)$$

Letting  $\varepsilon \rightarrow 0$  we get that  $x_0x_1 + x_0 = 0$ . It is possible to come to the same conclusion directly from (2.8). Namely, since there is no  $O(\varepsilon)$  term on the right-hand side of (2.8), then the coefficient of the  $O(\varepsilon)$  term on the left-hand side must be zero. This gives us the following problem.

$$O(\varepsilon) \quad x_0x_1 + x_0 = 0$$

The solution is  $x_1 = -1$ .

We have determined the first two terms in the expansion, but we could easily continue and find more. For example, since there is no  $O(\varepsilon^2)$  term on the right-hand side of (2.8), then the coefficient of the  $O(\varepsilon^2)$  term on the left-hand side must be zero. This gives us:

$$O(\varepsilon^2) \quad 2x_0x_2 + x_1^2 + 2x_1 = 0$$

The solutions are  $x_2 = \pm \frac{1}{2}$ .

The conclusion we make from the above procedure is that one of the solutions is

$$x \sim 1 - \varepsilon + \frac{1}{2}\varepsilon^2, \quad (2.10)$$

and the other is

$$x \sim -1 - \varepsilon - \frac{1}{2}\varepsilon^2. \quad (2.11)$$

These approximations hold for small  $\varepsilon$ , and for this reason they are said to be *asymptotic expansions* of the solutions as  $\varepsilon \rightarrow 0$ . As you might have noticed, the above expansions are the same as the ones we obtained directly from the formula for the solution, as given in (2.4).

## 2.2 How to Find a Regular Expansion

The ideas used to construct asymptotic expansions of the solutions of a quadratic equation are easily extended to more complex problems. Exactly how one proceeds depends on how the problem is stated, and the following three situations are the most common.

### 2.2.1 Given a Specific Function

The expansion in (2.3) is an example of this situation. For these problems Taylor's theorem, either directly or indirectly, is most often used to construct the expansion. In fact, it is not unusual to have to use it more than once.

*Example 1*

$$f(\varepsilon) = \sin\left(\frac{\pi}{6} + \varepsilon\right).$$

One way to find the expansion is to note that  $f(\varepsilon)$  is a smooth function of  $\varepsilon$ . So, it should be possible to use a Taylor series. Since  $f'(\varepsilon) = \cos\left(\frac{\pi}{6} + \varepsilon\right)$  and  $f''(\varepsilon) = -\sin\left(\frac{\pi}{6} + \varepsilon\right)$ , then

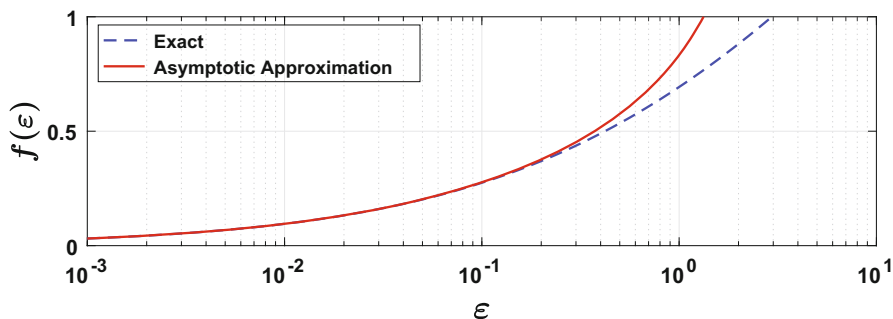
$$\begin{aligned}\sin\left(\frac{\pi}{6} + \varepsilon\right) &= f(0) + \varepsilon f'(0) + \frac{1}{2}\varepsilon^2 f''(0) + \cdots \\ &= \frac{1}{2} + \frac{1}{2}\sqrt{3}\varepsilon - \frac{1}{4}\varepsilon^2 + \cdots.\end{aligned}$$

With this we have a three term expansion of the function. ■

*Example 2*

$$f(\varepsilon) = \ln(1 + \sqrt{\varepsilon}).$$

The direct application of Taylor's theorem is not possible here as the function has a singularity at  $\varepsilon = 0$  (e.g.,  $f'(0)$  is not defined). However, setting  $y = \sqrt{\varepsilon}$ , then  $y = 0$  is equivalent to  $\varepsilon = 0$ , and  $\ln(1 + y)$  is well defined (and smooth) near  $y = 0$ .



**Fig. 2.2** The function  $f(\varepsilon) = \ln(1 + \sqrt{\varepsilon})$ , and its asymptotic approximation given in (2.12)

Since (see Table 2.1),  $\ln(1 + y) = y - \frac{1}{2}y^2 + \frac{1}{3}y^3 + \dots$  we obtain the three term expansion

$$\ln(1 + \sqrt{\varepsilon}) \sim \sqrt{\varepsilon} - \frac{1}{2}\varepsilon + \frac{1}{3}\varepsilon^{3/2}. \quad (2.12)$$

The accuracy of this approximation is shown in Fig. 2.2. As required, the approximation improves as  $\varepsilon$  gets closer to zero, with the two curves indistinguishable once  $\varepsilon$  drops below about  $10^{-1}$ . For those who might be interested, a proof that this is a valid asymptotic expansion is given in Exercise 2.18. ■

The above example involves a substitution, where a question about  $\varepsilon = 0$  is replaced with one about  $y = 0$ . If you are unfamiliar with this idea, it is recommended that you review the examples in Sect. A.1.1 to learn more about how substitution can be used with Taylor's theorem. As you will see, substitution plays a key role in several of the examples, and exercises, in this chapter.

### Example 3

$$f(\varepsilon) = \left[ \frac{1}{1 - \cos(\varepsilon)} \right]^3.$$

This function is not defined at  $\varepsilon = 0$ , so the straightforward Taylor series approach won't work. To find a two-term expansion of this for small  $\varepsilon$ , we start with the inner-most function, which is  $\cos(\varepsilon)$ . Using the Taylor expansion of  $\cos(x)$  given in Table 2.1, we have

$$\cos(\varepsilon) \sim 1 - \frac{1}{2}\varepsilon^2 + \frac{1}{24}\varepsilon^4 + \dots$$



With this

$$\begin{aligned}\frac{1}{1 - \cos(\varepsilon)} &\sim \frac{1}{\frac{1}{2}\varepsilon^2 - \frac{1}{24}\varepsilon^4 + \dots} \\ &= \frac{2}{\varepsilon^2} \frac{1}{1 - \frac{1}{12}\varepsilon^2 + \dots}.\end{aligned}$$

The significance of the last expression is that we have factored the function into a singular function, and one that is well behaved, near  $\varepsilon = 0$ . For the well-behaved factor, note that

$$\left[ \frac{1}{1 - \frac{1}{12}\varepsilon^2 + \dots} \right]^3 = (1 + y)^{-3},$$

where  $y = -\frac{1}{12}\varepsilon^2 + \dots$ . The assumption that  $\varepsilon$  is close to zero means that  $y$  is close to zero, and this means we can use the binomial expansion, which is the second entry in Table 2.1. In particular, with  $a = 1$  and  $\gamma = -3$ ,

$$\begin{aligned}\left[ \frac{1}{1 - \frac{1}{12}\varepsilon^2 + \dots} \right]^3 &= 1 - 3y + 6y^2 + \dots \\ &= 1 - 3\left(-\frac{1}{12}\varepsilon^2 + \dots\right) + 6\left(-\frac{1}{12}\varepsilon^2 + \dots\right)^2 + \dots \\ &= 1 + \frac{1}{4}\varepsilon^2 + \dots.\end{aligned}$$

The resulting two-term expansion is

$$\left[ \frac{1}{1 - \cos(\varepsilon)} \right]^3 \sim \frac{8}{\varepsilon^6} \left( 1 + \frac{1}{4}\varepsilon^2 \right). \blacksquare \quad (2.13)$$

Generalizing the above examples, the asymptotic expansions we derived have the form

$$f \sim f_0\varepsilon^\alpha + f_1\varepsilon^\beta + f_2\varepsilon^\gamma + \dots, \quad (2.14)$$

where  $\alpha < \beta < \gamma < \dots$  and the  $f_i$ 's are nonzero. Unlike a Taylor series, the exponents do not need to be positive integers. They must, however, satisfy  $\alpha < \beta < \gamma < \dots$ , which is known as a *well-ordering condition*. This is required because (2.14) is an approximation of  $f$  as  $\varepsilon \rightarrow 0^+$ .

Another way of writing the well-ordered requirement of  $\alpha < \beta$  is that

$$\lim_{\varepsilon \rightarrow 0^+} \frac{\varepsilon^\beta}{\varepsilon^\alpha} = 0. \quad (2.15)$$

Because well ordering plays such an important role in constructing an asymptotic expansion, it is convenient to introduce the  $\ll$  symbol. To state that

$$\varepsilon^\beta \ll \varepsilon^\alpha$$

means that (2.15) holds. As a special case of this, to state that  $\varepsilon^\beta \ll 1$  means that  $\beta > 0$ . The use of limits to formally define what it means to be an asymptotic expansion is considered in Sect. 2.3.

The question arises as to how many terms of an asymptotic expansion to determine. The answer depends on the values of  $\varepsilon$  under consideration, but for most applications the answer is no more than two or three. In fact, it is typical for the types of problems considered in Sects. 2.5 and 2.6 to only find the first term in the expansion.

## 2.2.2 Given an Algebraic or Transcendental Equation

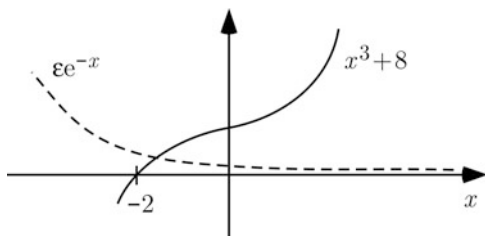
The idea here is that we are given an algebraic or transcendental equation and we want to construct an approximation of the solution(s). This is exactly what we did for the quadratic equation example (2.1).

*Example 1* We will find a two-term expansion of the solution(s) of

$$x^3 - \varepsilon e^{-x} + 8 = 0. \quad (2.16)$$

Although not required, it is strongly recommended that you first try to assess how many solutions there are and, if possible, their approximate location. The reason is that we will have to guess the form of the expansion and any information we might have about the solution(s) can be helpful. With this in mind, the equation is written as  $x^3 + 8 = \varepsilon e^{-x}$ , and the functions involved in this equation are sketched in Fig. 2.3. It shows that there is one real-valued solution that is located slightly to the right of  $x = -2$ .

**Fig. 2.3** Sketch of the functions appearing in the transcendental equation in (2.16)



The first step is to guess the form of the expansion. For many problems it can be difficult to know what to assume, and in such cases the usual assumption is that a power series is appropriate. What this means here is that we are going to assume that

$$x \sim x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \cdots. \quad (2.17)$$

This is going to be substituted into (2.16) and this requires us to expand  $e^x$ . To do this, note that

$$\begin{aligned} e^{-x} &\sim e^{-(x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \cdots)} \\ &= e^{-x_0} e^{-(\varepsilon x_1 + \varepsilon^2 x_2 + \cdots)}. \end{aligned}$$

Setting  $y = \varepsilon x_1 + \varepsilon^2 x_2 + \cdots$ , and noting that  $y$  is close to zero for small  $\varepsilon$ , then

$$\begin{aligned} e^{-x} &\sim e^{-x_0} e^{-y} \\ &= e^{-x_0} \left( 1 - y + \frac{1}{2} y^2 + \cdots \right) \\ &= e^{-x_0} \left( 1 - (\varepsilon x_1 + \varepsilon^2 x_2 + \cdots) + \frac{1}{2} (\varepsilon x_1 + \varepsilon^2 x_2 + \cdots)^2 + \cdots \right) \\ &= e^{-x_0} (1 - \varepsilon x_1 + \cdots). \end{aligned}$$

Using the binomial expansion, given in Table 2.1, and (2.17) we also have that

$$x^3 \sim x_0^3 + 3\varepsilon x_0^2 x_1 + \cdots.$$

With this, the original equation given in (2.16) takes the form

$$x_0^3 + 3\varepsilon x_0^2 x_1 + \cdots - \varepsilon e^{-x_0} (1 - \varepsilon x_1 + \cdots) + 8 = 0. \quad (2.18)$$

The first problem to solve is obtained by simply setting  $\varepsilon = 0$ , which gives us the following  $O(1)$  problem.

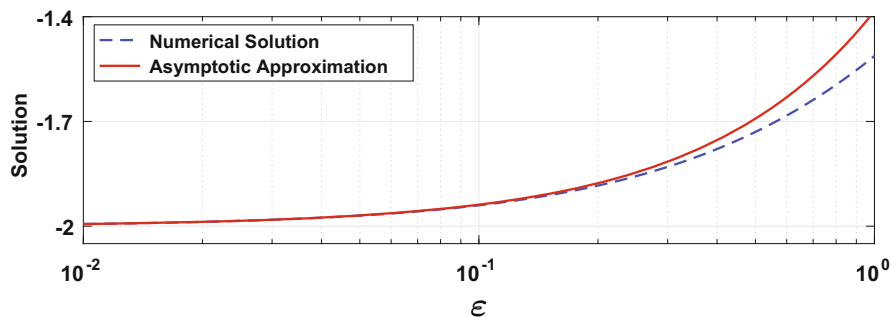
$$O(1) \quad x_0^3 + 8 = 0$$

The real-valued solution is  $x_0 = -2$ .

As explained when working with (2.8), the next problem is obtained by noting that the combined coefficient of the  $O(\varepsilon)$  terms on the left must add to zero.

$$O(\varepsilon) \quad 3x_0^2 x_1 - e^{-x_0} = 0$$

The solution is  $x_1 = \frac{1}{3x_0^2} e^{-x_0} = \frac{1}{12} e^2$ .



**Fig. 2.4** Comparison between the numerical solution of (2.16) and the asymptotic expansion (2.19) as a function of  $\varepsilon$

We have therefore found that a two-term expansion of the solution is

$$x \sim -2 + \frac{1}{12}\varepsilon e^2. \quad (2.19)$$

This expansion is plotted in Fig. 2.4 along with the numerical solution. The asymptotic nature of the approximation is evident as  $\varepsilon \rightarrow 0$ . ■

*Example 2* The equation to solve is

$$x^3 - \varepsilon e^{-x} = 0. \quad (2.20)$$

Writing this as  $x^3 = \varepsilon e^{-x}$ , the functions in this equation are sketched in Fig. 2.5. It is apparent that there is one real-valued solution, and it approaches  $x = 0$  as  $\varepsilon$  decreases to zero. Our goal is to derive a two-term approximation of this solution.

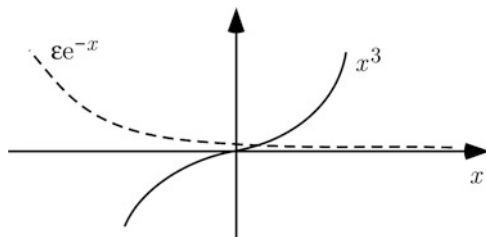
The first point to make is how similar this equation is to the one in the last example. What is surprising is that the expansion in (2.17) will not work here. One indication of this is that we know that the solution approaches zero as  $\varepsilon$  decreases to zero, but (2.17) assumes it approaches  $x_0$ . More significantly, if  $x \approx 0$ , then  $e^{-x} \approx 1$ , and the equation is, approximately,  $x^3 = \varepsilon$ . This indicates that  $x \approx \varepsilon^{1/3}$ , and, if so, then (2.17) will certainly fail.

What is needed is the general version of an expansion as given in (2.14). This means that we are assuming

$$x \sim x_0 \varepsilon^\alpha + x_1 \varepsilon^\beta + x_2 \varepsilon^\gamma + \dots, \quad (2.21)$$

where  $\alpha < \beta < \gamma < \dots$ . In fact, since we know that  $x$  goes to zero with  $\varepsilon$ , we are able to state that  $0 < \alpha < \beta < \gamma < \dots$ . Finally, it is assumed that the  $x_i$ 's are nonzero.

In preparation for substituting (2.21) into the equation, note that (see Exercise 2.38)



**Fig. 2.5** Sketch of the functions appearing in the transcendental equation in (2.20)

$$\begin{aligned} x^3 &\sim (x_0 \varepsilon^\alpha + x_1 \varepsilon^\beta + x_2 \varepsilon^\gamma + \dots)^3 \\ &= x_0^3 \varepsilon^{3\alpha} + 3x_0^2 x_1 \varepsilon^{2\alpha+\beta} + \dots \end{aligned}$$

Also, since  $x$  is close to zero

$$\begin{aligned} e^{-x} &= 1 - x + \frac{1}{2}x^2 + \dots \\ &= 1 - x_0 \varepsilon^\alpha + \dots \end{aligned}$$

Equation (2.20) now becomes

$$\begin{array}{ccccccc} x_0^3 \varepsilon^{3\alpha} & + & 3x_0^2 x_1 \varepsilon^{2\alpha+\beta} & + & \dots & - & \varepsilon(1 - x_0 \varepsilon^\alpha + \dots) = 0. \\ \textcircled{1} & & \textcircled{2} & & & & \textcircled{3} \quad \textcircled{4} \end{array} \quad (2.22)$$

Our first task is to determine  $\alpha$ , and this is done using a process called *balancing*. To explain, according to the above equation, the terms on the left add to zero. We can, if needed, get this to happen by simply taking the  $x_i$ 's equal to zero. The exception is the term labeled with a ③. This means that at least one of the other terms must balance with it, so they sum to zero. Given the requirement that  $0 < \alpha < \beta$ , and that  $x_0 \neq 0$ , it follows that the only term capable of balancing with ③ is ①. For this to happen, the exponents of these terms must agree, and this means that we need  $3\alpha = 1$ . From this we conclude that  $\alpha = 1/3$ .

$$O(\varepsilon^{1/3}) \quad x_0^3 - 1 = 0$$

The solution is  $x_0 = 1$ .

Our next task is to use balancing to determine  $\beta$ . Since  $x_0$  is nonzero, we are left with the term labeled with a ④ in (2.22). The only term available to balance with it is ②. For this to happen, the  $\varepsilon$  exponents of ② and ④ must agree, and so we need  $2\alpha + \beta = 1 + \alpha$ . In other words,  $\beta = 2/3$ .

$$O(\varepsilon^{4/3}) \quad 3x_0^2 x_1 + x_0 = 0$$

The solution is  $x_1 = -1/3$ .

We have therefore found that a two-term expansion of the solution is

$$x \sim \varepsilon^{1/3} - \frac{1}{3}\varepsilon^{2/3}. \blacksquare \quad (2.23)$$

### 2.2.3 Given an Initial Value Problem

The next stage in the development is to apply regular expansions to problems involving differential equations. We will work out two examples, the first involves a single equation, and the second a system.

*Example 1* The projectile problem furnishes an excellent example. Using (1.75)–(1.77) the problem to solve is

$$\frac{d^2u}{d\tau^2} = -\frac{1}{(1 + \varepsilon u)^2}, \quad (2.24)$$

where

$$u(0) = 0, \quad (2.25)$$

$$\frac{du}{d\tau}(0) = 1. \quad (2.26)$$

It is important to note that we are using the nondimensional problem and not the original given in (1.1)–(1.3). The assumption used in the scaling is that the initial velocity  $v_0$  is small, which lead us to the definition of  $\varepsilon$  given in (1.78). Consequently, what we are interested in here is an expansion of the solution for small  $\varepsilon$ .

The procedure for constructing the expansion will mimic what was done earlier. We start by stating what we believe to be the appropriate form for the expansion. A reasonable assumption, without knowing anything else about the solution, is that a simple power series expansion can be used. In other words, our assumption is

$$u \sim u_0(\tau) + \varepsilon u_1(\tau) + \cdots. \quad (2.27)$$

The expansion is suppose to identify how the solution depends on  $\varepsilon$ . The terms in the expansion can, and almost inevitability will, depend on the other variables and parameters in the problem. For the projectile problem this means that each term in the expansion depends on (nondimensional) time and this dependence is included in (2.27).

In preparation for substituting (2.27) into (2.24) note

$$\begin{aligned}\frac{1}{(1 + \varepsilon u)^2} &= 1 - 2\varepsilon u + 3\varepsilon^2 u^2 + \cdots \\ &\sim 1 - 2\varepsilon(u_0 + \varepsilon u_1 + \cdots) + 3\varepsilon^2(u_0 + \cdots)^2 + \cdots \\ &= 1 - 2\varepsilon u_0 + \cdots.\end{aligned}$$

With this, the differential equation (2.24) becomes

$$u_0'' + \varepsilon u_1'' + \cdots = -1 + 2\varepsilon u_0 + \cdots. \quad (2.28)$$

It is critical that the initial conditions are also included, and for these we have

$$\begin{aligned}u_0(0) + \varepsilon u_1(0) + \cdots &= 0, \\ u_0'(0) + \varepsilon u_1'(0) + \cdots &= 1.\end{aligned}$$

As usual we break the above equations down into problems depending on the power of  $\varepsilon$ .

$$\begin{aligned}O(1) \quad u_0'' &= -1 \\ u_0(0) &= 0, u_0'(0) = 1\end{aligned}$$

The solution of this problem is  $u_0 = \tau(1 - \frac{1}{2}\tau)$ .

$$\begin{aligned}O(\varepsilon) \quad u_1'' &= 2u_0 \\ u_1(0) &= 0, u_1'(0) = 0\end{aligned}$$

The solution of this problem is  $u_1 = \frac{1}{12}\tau^3(4 - \tau)$ .

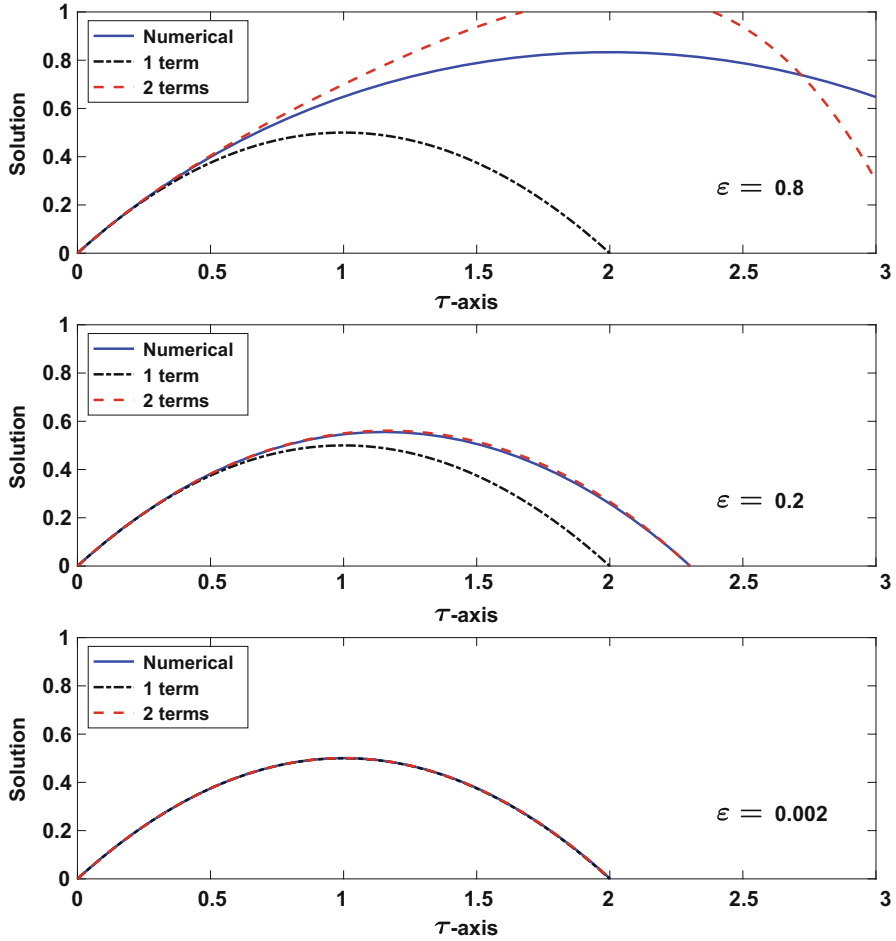
We have therefore found that a two-term expansion of the solution is

$$u \sim \frac{1}{2}\tau(2 - \tau) + \frac{1}{12}\varepsilon\tau^3(4 - \tau). \quad (2.29)$$

To determine how well we have done in approximating the solution, a comparison with the numerical solution is shown in Fig. 2.6 for three values of  $\varepsilon$ . It is seen that the one-term approximation,  $u \sim \tau(2 - \tau)/2$ , produces a reasonably accurate approximation for  $\varepsilon = 0.002$ , but not when  $\varepsilon = 0.2$  or  $\varepsilon = 0.8$ . In contrast, the two-term approximation (2.29) does well for  $\varepsilon = 0.002$  and  $\varepsilon = 0.2$  but not when  $\varepsilon = 0.8$ . The fact that the expansion is inaccurate when  $\varepsilon = 0.8$  is not surprising as the derivation is based on the assumption that  $\varepsilon$  is relatively small.

The expansion can be written in dimensional variables by using the scaling introduced in Sect. 1.5.1. Specifically,  $x = x_c u$  and  $t = t_c \tau$ , where  $x_c = v_0^2/g$  and  $t_c = v_0/g$ . In this case, from (2.29),

$$x \sim \frac{1}{2}t(2v_0 - gt) + \frac{g}{12R}t^3(4v_0 - gt). \quad (2.30)$$



**Fig. 2.6** Comparison between the numerical solution of the projectile problem (2.24)–(2.26), and the asymptotic expansion (2.29), for increasing values of  $\varepsilon$

Note that the first term, which comes from  $u_0$ , is  $t(2v_0 - gt)/2$ , and the second term, which comes from  $\varepsilon u_1$ , is  $gt^3(4v_0 - gt)/(12R)$ . Physically, the first term gives the position of the projectile for a uniform gravitational field, which we originally derived in Sect. 1.1. The second term is the correction due to the nonlinear gravitational field.

It is worth expressing the conclusions of this example in physical terms. To do this, recall that the speed of sound is about 343 m/s. So, since  $\varepsilon = v_0/(gR)$ , then for initial velocities up to about Mach 2 the uniform gravitational field approximation is reasonably accurate. In comparison, the corrected approximation, which is given in (2.30), is accurate for initial velocities up to about Mach 10. Although this is a



rather large velocity, it is still quite a bit smaller than the escape velocity, which is about Mach 33. ■

*Example 2* The ideas used to find an approximation for a single equation are easily extended to systems. As an example, consider the thermokinetic model of Exercise 1.24. In nondimensional variables, the equations are

$$\frac{du}{dt} = 1 - ue^{\varepsilon(q-1)}, \quad (2.31)$$

$$\frac{dq}{dt} = ue^{\varepsilon(q-1)} - q. \quad (2.32)$$

The initial conditions are  $u(0) = q(0) = 0$ . We are assuming here that the nonlinearity is weak, which means that  $\varepsilon$  is small. Also, to simplify the problem, the other parameters that appear in the nondimensionalization have been set to one.

Generalizing (2.27), we expand both functions using our usual assumption, which is that

$$u \sim u_0(t) + \varepsilon u_1(t) + \cdots,$$

$$q \sim q_0(t) + \varepsilon q_1(t) + \cdots.$$

Before substituting these into the differential equations, note that

$$\begin{aligned} e^{\varepsilon(q-1)} &\sim 1 + \varepsilon(q-1) + \frac{1}{2}\varepsilon^2(q-1)^2 + \cdots \\ &\sim 1 + \varepsilon(q_0 + \varepsilon q_1 + \cdots - 1) + \frac{1}{2}\varepsilon^2(q_0 + \varepsilon q_1 + \cdots - 1)^2 + \cdots \\ &\sim 1 + \varepsilon(q_0 - 1) + \cdots, \end{aligned}$$

and

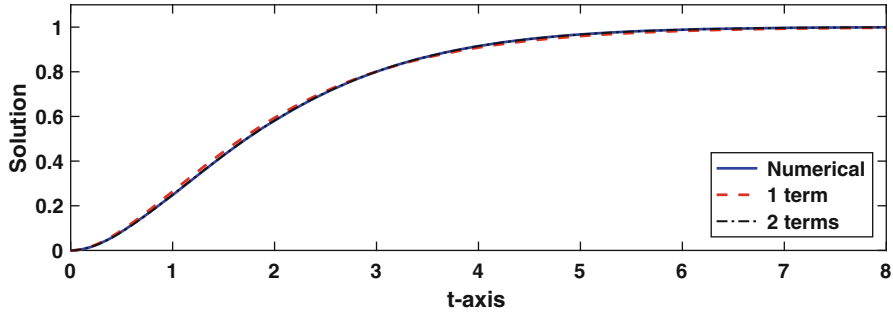
$$\begin{aligned} ue^{\varepsilon(q-1)} &\sim (u_0 + \varepsilon u_1 + \cdots)[1 + \varepsilon(q_0 - 1) + \cdots] \\ &\sim u_0 + \varepsilon[u_0(q_0 - 1) + u_1] + \cdots. \end{aligned}$$

With this, (2.31) and (2.32) take the form

$$\begin{aligned} u'_0 + \varepsilon u'_1 + \cdots &= 1 - u_0 - \varepsilon(u_0(q_0 - 1) + u_1) + \cdots, \\ q'_0 + \varepsilon q'_1 + \cdots &= u_0 - q_0 + \varepsilon(u_0(q_0 - 1) + u_1 - q_1) + \cdots. \end{aligned}$$

As usual we break the above equations down into problems depending on the power of  $\varepsilon$ .

$$\begin{aligned} O(1) \quad u'_0 &= 1 - u_0 \\ q'_0 &= u_0 - q_0 \end{aligned}$$



**Fig. 2.7** Comparison between the numerical solution for  $q(t)$ , and the asymptotic expansion (2.34). In the calculation,  $\varepsilon = 0.1$

The general solution of the first equation is  $u_0 = 1 - Ae^{-t}$ . Since  $u_0(0) = 0$ , it follows that  $u_0 = 1 - e^{-t}$ . Substituting this into the second equation and solving one finds that  $q_0 = 1 - (1 + t)e^{-t}$ .

$$O(\varepsilon) \quad \begin{aligned} u_1' &= -u_1 - u_0(q_0 - 1) \\ q_1' &= -q_1 + u_1 + u_0(q_0 - 1) \end{aligned}$$

The initial conditions are  $u_1(0) = q_1(0) = 0$ . The equation for  $u_1$  is first order, and the solution can be found using an integrating factor. Once  $u_1$  is determined, then the  $q_1$  equation can be solved using an integrating factor. Carrying out the calculation one finds that  $u_1 = \frac{1}{2}(t^2 + 2t - 4)e^{-t} + (2 + t)e^{-2t}$ ,  $q_1 = \frac{1}{6}(t^3 - 18t + 30)e^{-t} - (2t + 5)e^{-2t}$ .

We have therefore found that a two-term expansion of the solution is

$$u(t) \sim 1 - e^{-t} + \varepsilon \left( \frac{1}{2}(t^2 + 2t - 4)e^{-t} + (2 + t)e^{-2t} \right), \quad (2.33)$$

$$q(t) \sim 1 - (1 + t)e^{-t} + \varepsilon \left( \frac{1}{6}(t^3 - 18t + 30)e^{-t} - (2t + 5)e^{-2t} \right). \quad (2.34)$$

A comparison of the numerical solution for  $q(t)$ , and the above asymptotic approximation for  $q(t)$  is shown in Fig. 2.7 for  $\varepsilon = 0.1$ . It is seen that even the one-term approximation,  $q \sim 1 - (1 + t)e^{-t}$ , produces a reasonably accurate approximation, while the two-term approximation is indistinguishable from the numerical solution. The approximations for  $u(t)$ , which are not shown, are also as accurate. ■

## 2.3 Scales and Approximation

All but three of the asymptotic expansions derived in the previous sections have the form of a power series in  $\varepsilon$ . The exceptions are (2.12), (2.13), and (2.23). To include them, we can state that all of the expansions have the form

$$f \sim f_1 \varepsilon^\alpha + f_2 \varepsilon^\beta + f_3 \varepsilon^\gamma + \dots, \quad (2.35)$$

where  $\alpha < \beta < \gamma < \dots$ . This gives rise to the introduction of scale functions, which in the above expression are  $\phi_1 = \varepsilon^\alpha$ ,  $\phi_2 = \varepsilon^\beta$ ,  $\phi_3 = \varepsilon^\gamma$ ,  $\dots$ . The reason they qualify as scale functions is that

$$\lim_{\varepsilon \rightarrow 0^+} \frac{\phi_{i+1}}{\phi_i} = 0, \text{ for } i = 1, 2, 3, \dots \quad (2.36)$$

More generally,  $\phi_1, \phi_2, \phi_3, \dots, \phi_n$  are *scale functions*, and can be used to construct an asymptotic expansion, if the above limit holds for  $i = 1, 2, \dots, n - 1$ . Said another way, the limit in (2.36) means that the  $\phi_i$ 's are *well ordered*.

Because of the importance of being well ordered, it is convenient to have a mathematical way to express this requirement. This is done using the  $\ll$  symbol introduced earlier. In this text, to state that  $\phi(\varepsilon) \ll \varphi(\varepsilon)$  will mean that

$$\lim_{\varepsilon \rightarrow 0^+} \frac{\phi(\varepsilon)}{\varphi(\varepsilon)} = 0.$$

So, for example,  $\varepsilon^2 \ll 1 + \varepsilon$ ,  $\varepsilon \ll e^\varepsilon$ , and  $\sin \varepsilon \ll \cos \varepsilon$ .

The scale functions most often used come from Taylor's theorem, so

$$1. \quad \phi_1 = 1, \phi_2 = \varepsilon, \phi_3 = \varepsilon^2, \dots$$

The more general power functions

$$2. \quad \phi_1 = \varepsilon^\alpha, \phi_2 = \varepsilon^\beta, \phi_3 = \varepsilon^\gamma, \dots \text{ where } \alpha < \beta < \gamma < \dots$$

are also common, as demonstrated in (2.13) and (2.23). A third set of scale functions often used involve an exponential dependence, and an example is

$$3. \quad \phi_1 = 1, \phi_2 = e^{-1/\varepsilon}, \phi_3 = e^{-2/\varepsilon}, \dots$$

These will make an appearance later, in Sect. 2.5, when we examine what are called boundary layers.

The reason the limit in (2.36) is important is that it plays a central role in the definition of an asymptotic expansion. Namely, suppose we have a set of scale functions  $\phi_1, \phi_2, \phi_3, \dots$ . In this case,  $f \sim f_1 \phi_1$  is a one term asymptotic expansion if

$$\lim_{\varepsilon \rightarrow 0^+} \frac{f - f_1 \phi_1}{\phi_1} = 0. \quad (2.37)$$

Similarly,  $f \sim f_1 \phi_1 + f_2 \phi_2$  is a two term asymptotic expansion if the above limit holds, and

$$\lim_{\varepsilon \rightarrow 0^+} \frac{f - f_1 \phi_1 - f_2 \phi_2}{\phi_2} = 0. \quad (2.38)$$

In each case, the limit means that the error in the approximation (the numerator) goes to zero faster than the last scale function used in the approximation. As demonstrated in the examples in the previous section, the above limits are not explicitly used to find an expansion. However, they are useful for those interested in the theoretical foundations of the subject. For us, the critical point is that the asymptotic expansion is determined by how the function, or solution, behaves as  $\varepsilon \rightarrow 0$ .

It is possible to use the above limits to help explain why exponential scale functions are sometimes needed. The reason is that the power functions are not able to describe exponential behavior. To illustrate, suppose it is assumed that  $e^{-1/\varepsilon} \sim x_0 + \varepsilon x_1 + \varepsilon^2 x_2 + \dots$ . In this case, the coefficients must satisfy the following limits:

$$\begin{aligned} x_0 &= \lim_{\varepsilon \rightarrow 0^+} e^{-1/\varepsilon}, \\ x_1 &= \lim_{\varepsilon \rightarrow 0^+} \frac{e^{-1/\varepsilon} - x_0}{\varepsilon}, \\ x_2 &= \lim_{\varepsilon \rightarrow 0^+} \frac{e^{-1/\varepsilon} - x_0 - \varepsilon x_1}{\varepsilon^2}. \end{aligned}$$

Using l'Hospital's rule one finds that each limit is zero, and so  $x_0 = 0$ ,  $x_1 = 0$ ,  $x_2 = 0$ ,  $\dots$ . In other words, as far as the functions  $1$ ,  $\varepsilon$ ,  $\varepsilon^2$ ,  $\dots$  are concerned,  $e^{-1/\varepsilon}$  is just zero. This function certainly has rather small values but it is not identically zero. What is happening is that  $e^{-1/\varepsilon}$  goes to zero so quickly that the power functions are not able to describe it other than to just conclude that the function is zero. In this case  $e^{-1/\varepsilon}$  is said to be *transcendently small* relative to the power functions.

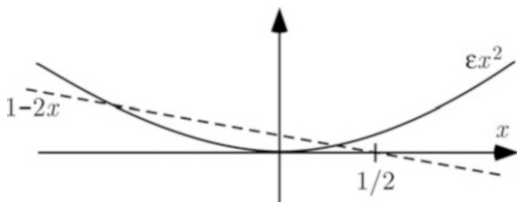
It is also worth pointing out that the definition of an asymptotic expansion does not say anything about what happens when more terms are used in the expansion for a given value of  $\varepsilon$ . If we were to calculate every term in the expansion, and produce an infinite series in the process, the fact that it is an asymptotic expansion does not mean the series has to converge. In fact, some of the more interesting asymptotic expansions diverge. For this reason it is inappropriate to use an equal sign in (2.6) and why the symbol  $\sim$  is used instead.

## 2.4 Introduction to Singular Perturbations

All of the equations considered up to this point produce regular expansions. This means, roughly, that the expansions can be found without having to rescale the problem. We now turn our attention to those that are not regular, what are known as singular perturbation problems. The first example considered is the quadratic equation

$$\varepsilon x^2 + 2x - 1 = 0. \quad (2.39)$$

**Fig. 2.8** Sketch of the functions appearing in the quadratic equation in (2.39)



A tip-off that this is singular is that  $\varepsilon$  multiplies the highest-order term in the equation. Setting  $\varepsilon = 0$  drops the order down to linear, and this has dramatic effects on the number and types of solutions.

The best place to begin is to sketch the functions in the equation to get an idea of the number and location of the solutions. This is done in Fig. 2.8, which shows that there are two real-valued solutions. One is close to  $x = \frac{1}{2}$  and for this reason it is expected that an expansion of the form  $x \sim x_0 + x_1\varepsilon + \dots$  will work. In contrast, the location of the second solution looks to approach  $-\infty$  as  $\varepsilon$  gets close to zero. Consequently, we should not be shocked later when we find that the expansion for this solution has the form  $x \sim x_0\varepsilon^{-1} + \dots$ , where  $x_0$  is negative.

We start out as if this were a regular perturbation problem and assume the solutions can be expanded as

$$x \sim x_0 + \varepsilon x_1 + \dots \quad (2.40)$$

Substituting this into (2.39) we obtain

$$\varepsilon(x_0^2 + 2\varepsilon x_0 x_1 + \dots) + 2(x_0 + \varepsilon x_1 + \dots) - 1 = 0. \quad (2.41)$$

Equating like powers of  $\varepsilon$  produces the following problems.

$$O(1) \quad 2x_0 - 1 = 0$$

The solution is  $x_0 = \frac{1}{2}$ .

$$O(\varepsilon) \quad x_0^2 + 2x_1 = 0$$

The solution is  $x_1 = -\frac{1}{8}$ .

We therefore have

$$x \sim \frac{1}{2} - \frac{1}{8}\varepsilon. \quad (2.42)$$

This expansion is consistent with the conclusions we derived earlier from Fig. 2.8 for one of the solutions. It is also apparent that no matter how many terms we calculate in the expansion (2.40) we will not obtain the second solution.

The failure of the regular expansion to find all of the solutions is typical of a singular perturbation problem. The method used to remedy the situation is to introduce a scaling transformation. Specifically, we will change variables and let

$$\bar{x} = \frac{x}{\varepsilon^\gamma}. \quad (2.43)$$

With this, (2.39) takes the form

$$\varepsilon^{1+2\gamma} \bar{x}^2 + 2\varepsilon^\gamma \bar{x} - 1 = 0. \quad (2.44)$$

①                      ②                      ③

The reason for not finding two solutions earlier was that the quadratic term was lost when  $\varepsilon = 0$ . Given the fact that this term is why there are two solutions in the first place we need to determine how to keep it in the equation as  $\varepsilon \rightarrow 0$ . In other words, term ① in (2.44) must balance with one of the other terms and this must be the first problem solved as  $\varepsilon \rightarrow 0$ . For example, suppose we assume the balance is between terms ① and ③, while term ② is of higher or equal order. For this to occur, we need  $O(\varepsilon^{1+2\gamma}) = O(1)$  and this would mean  $\gamma = -\frac{1}{2}$ . With this ①, ③ =  $O(1)$  and ② =  $O(\varepsilon^{-1/2})$ . This result is inconsistent with our original assumption that ② is higher order. Therefore, the balance must be with another term. This type of argument is central to singular perturbation problems and we will use a table format to present the steps used to determine the correct balance (note that  $\ll$  is defined on page 65).

Balance	Condition on $\gamma$	Consistency check	Conclusion
① $\sim$ ③ with ② $\ll$ ①, ③	$1 + 2\gamma = 0$ $\Rightarrow \gamma = -1/2$	①, ③ = $O(\varepsilon)$ ② = $O(\varepsilon^{-1/2})$	Inconsistent with balance
① $\sim$ ② with ③ $\ll$ ①, ②	$1 + 2\gamma = \gamma$ $\Rightarrow \gamma = -1$	①, ② = $O(\varepsilon^{-1})$ ③ = $O(1)$	Consistent with balance

Based on the above analysis,  $\gamma = -1$  and with this the equation takes the form

$$\bar{x}^2 + 2\bar{x} - \varepsilon = 0. \quad (2.45)$$

With this we assume our usual expansion, which is

$$\bar{x} \sim \bar{x}_0 + \bar{x}_1 \varepsilon + \dots. \quad (2.46)$$

The equation in this case becomes

$$\bar{x}_0^2 + 2\bar{x}_0\bar{x}_1\varepsilon + \dots + 2(\bar{x}_0 + \bar{x}_1\varepsilon + \dots) - \varepsilon = 0. \quad (2.47)$$

This gives us the following problems.

$$O(1) \quad \bar{x}_0^2 + 2\bar{x}_0 = 0$$

The solutions are  $\bar{x}_0 = -2$  and  $\bar{x}_0 = 0$ .

$$O(\varepsilon) \quad 2\bar{x}_0\bar{x}_1 + 2\bar{x}_1 - 1 = 0$$

If  $\bar{x}_0 = -2$ , then  $\bar{x}_1 = -\frac{1}{2}$ , while if  $\bar{x}_1 = 0$ , then  $\bar{x}_1 = \frac{1}{2}$ .

It might appear that we have somehow produced three solutions, the one in (2.42) along with the two found above. However, it is not hard to show that the solution corresponding to  $\bar{x}_0 = 0$  is the same one that was found earlier using a regular expansion. Consequently, the sought-after second solution is

$$x \sim \frac{1}{\varepsilon} \left( -2 - \frac{1}{2}\varepsilon \right). \quad (2.48)$$

The procedure used to derive this result contains many of the ideas we used to find regular expansions. The most significant difference is the introduction of a scaled variable, (2.43), and the subsequent balancing used to determine how the highest-order term participates in the problem. As we will see shortly, these will play a critical role when analyzing similar problems involving differential equations.

## 2.5 Introduction to Boundary Layers

As our introductory example of a singular perturbation problem involving a differential equation we will consider solving

$$\varepsilon y'' + 2y' + 2y = 0, \quad \text{for } 0 < x < 1, \quad (2.49)$$

where the boundary conditions are

$$y(0) = 0, \quad (2.50)$$

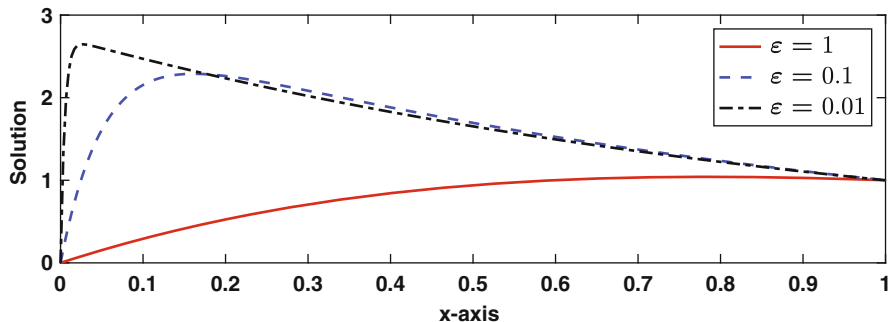
and

$$y(1) = 1. \quad (2.51)$$

This is a boundary value problem and it has the telltale signs of a singular perturbation problem. Namely,  $\varepsilon$  is multiplying the highest derivative so setting  $\varepsilon = 0$  results in a lower order problem.

This problem has been selected as the introductory problem because it can be solved exactly, and we will be able to use this to evaluate the accuracy of our approximations. To find the exact solution one assumes  $y = e^{rx}$  and from the differential equation concludes that  $r_{\pm} = (-1 \pm \sqrt{1 - 2\varepsilon})/\varepsilon$ . With this the general solution is  $y = c_1 e^{r_+ x} + c_2 e^{r_- x}$ . Imposing the two boundary conditions one finds that

$$y = \frac{e^{r_+ x} - e^{r_- x}}{e^{r_+} - e^{r_-}}. \quad (2.52)$$



**Fig. 2.9** Graph of the exact solution of the boundary value problem (2.49)–(2.50), for various values of  $\varepsilon$ . Note the appearance of the boundary layer near  $x = 0$  as  $\varepsilon$  decreases

This function is plotted in Fig. 2.9, for various values of  $\varepsilon$ . It is seen that as  $\varepsilon$  decreases the solution starts to show a rapid transition in the region near  $x = 0$ . Also, you will notice that the rapid change takes place over a spatial interval that has a width about equal to the size of  $\varepsilon$ . It is this region that forms the *boundary layer*. The interval outside this region forms what we will call the *outer region*.

It is typical in physical problems to know where the boundary layers are located. Consequently, in the examples and exercises in this textbook it will be stated where the layer is, and the task will then be to derive an asymptotic approximation for the solution.

The objective is, given that there is a boundary layer at  $x = 0$ , find a first term approximation of the solution over the interval  $0 \leq x \leq 1$ . This will be done by finding approximations in the outer and boundary layer regions, and then connecting them to form an approximation over the entire interval.

### Step 1: Outer Solution

The first step is simply to use a regular expansion and see what results. Similar to what we did with the projectile problem, it is assumed that

$$y \sim y_0(x) + \varepsilon y_1(x) + \cdots. \quad (2.53)$$

Introducing this into (2.49) we obtain

$$\varepsilon(y_0'' + \varepsilon y_1'' + \cdots) + 2(y_0' + \varepsilon y_1' + \cdots) + 2(y_0 + \varepsilon y_1 + \cdots) = 0, \quad (2.54)$$

where

$$y_0(1) + \varepsilon y_1(1) + \cdots = 1. \quad (2.55)$$

Only the boundary condition at  $x = 1$  has been included here. The reason is that we are building an approximation of the solution outside of the boundary layer, and so



it is incorrect to assume it can satisfy the condition at  $x = 0$ . With this, we obtain the following problems.

$$O(1) \quad 2y'_0 + 2y_0 = 0$$

$$y_0(1) = 1$$

The general solution of the differential equation is  $y_0 = ae^{-x}$ , where  $a$  is an arbitrary constant. From the boundary condition one finds that  $a = e$ , and so  $y_0(x) = e^{1-x}$ .

$$O(\varepsilon) \quad y''_0 + 2y'_1 + 2y_1 = 0$$

$$y_1(1) = 0$$

The general solution of the differential equation is  $y_1 = (b - x/2)e^{1-x}$ , where  $b$  is an arbitrary constant. With the given boundary condition we obtain  $y_1(x) = (1 - x)e^{1-x}/2$ .

Our regular expansion has yielded

$$y \sim e^{1-x} + \dots \quad (2.56)$$

Only the first term has been included here as this is all we are going to determine. The second term was found just to demonstrate that it is easy to find the other terms in the expansion if required.

### Step 2: Boundary Layer Solution

We will now construct an approximation of the solution in the neighborhood of  $x = 0$ , which corresponds to the interval where the function undergoes a rapid increase as shown in Fig. 2.9. Given its location, the approximation is called the *boundary layer solution*. It is also known as an inner solution, and correspondingly, the approximation in (2.56) is the *outer solution*. The width of this layer shrinks as  $\varepsilon \rightarrow 0$ , so we must make a change of variables to account for this. With this in mind we introduce the *boundary layer coordinate*

$$\bar{x} = \frac{x}{\varepsilon^\gamma} \quad (2.57)$$

The exact value of  $\gamma$  will be determined shortly but we already have some inkling what it might be. We saw in Fig. 2.9 that the rapid change in the solution near  $x = 0$  takes place over an interval that has width of about  $\varepsilon$ . So, it should not be too surprising that we will find that  $\gamma = 1$ . In any case, using the chain rule

$$\frac{d}{dx} = \frac{d\bar{x}}{dx} \frac{d}{d\bar{x}} = \frac{1}{\varepsilon^\gamma} \frac{d}{d\bar{x}} \quad (2.58)$$

and

$$\frac{d^2}{dx^2} = \frac{1}{\varepsilon^{2\gamma}} \frac{d^2}{d\bar{x}^2} \quad (2.59)$$

We will designate the solution as  $Y(\bar{x})$  when using  $\bar{x}$  as the independent variable. With this, the differential equation becomes

$$\varepsilon^{1-2\gamma} Y'' + 2\varepsilon^{-\gamma} Y' + 2Y = 0. \quad (2.60)$$

①                      ②                      ③.

We determine  $\gamma$  by balancing the terms in the above equation. Our goal is for the highest derivative to remain in the equation as  $\varepsilon \rightarrow 0$ . This gives us the following two possibilities (note that  $\ll$  is defined on page 65):

Balance		Condition on $\gamma$	Consistency Check	Conclusion
① $\sim$ ③ with ② $\ll$ ①, ③		$1 - 2\gamma = 0$ $\Rightarrow \gamma = 1/2$	①, ③ = $O(1)$ ② = $O(\varepsilon^{-1/2})$	Inconsistent with balance
① $\sim$ ② with ③ $\ll$ ①, ②		$1 - 2\gamma = -\gamma$ $\Rightarrow \gamma = 1$	①, ② = $O(\varepsilon^{-1})$ ③ = $O(1)$	Consistent with balance

Based on the above analysis we take  $\gamma = 1$  and with this the differential equation takes the form

$$Y'' + 2Y' + 2\varepsilon Y = 0. \quad (2.61)$$

Assuming  $Y(\bar{x}) \sim Y_0(\bar{x}) + \varepsilon Y_1(\bar{x}) + \dots$ , the differential equation becomes

$$(Y_0'' + \dots) + 2(Y_0' + \dots) + 2\varepsilon(Y_0 + \dots) = 0, \quad (2.62)$$

where, from (2.51),

$$Y_0(0) + \varepsilon Y_1(0) + \dots = 0.$$

Note that the boundary condition at  $x = 0$  has been included here but not the one at  $x = 1$ . The reason is that we are building an approximation of the solution in the immediate vicinity of  $x = 0$  and it is incorrect to assume it can satisfy the condition at the other end of the interval. With this, we obtain the following problem.

$$O(1) \quad Y_0'' + 2Y_0' = 0$$

$$Y_0(0) = 0$$

The general solution of the differential equation is  $Y_0 = A + Be^{-2\bar{x}}$ , where  $A$  and  $B$  are arbitrary constants. With the given boundary condition this reduces to  $Y_0 = A(1 - e^{-2\bar{x}})$ .

The approximation of the solution in the boundary layer is

$$Y(\bar{x}) \sim A(1 - e^{-2\bar{x}}) + \dots. \quad (2.63)$$

We will determine  $A$  by connecting this result with the approximation we have for the outer region, and this brings us to the next step.

### Step 3: Matching

We have made several assumptions about the solution and it is now time to prove that they are compatible. To explain what this means, our approximation consists of two different expansions, and each applies to a different part of the interval. The situation we find ourselves in is sketched in Fig. 2.10. This indicates that when coming out of the boundary layer the approximation in (2.63) approaches a constant value  $A$ . Similarly, the outer solution approaches a constant value,  $e$ , as it enters the boundary layer. There is a transition region, what is usually called an *overlap domain*, where the two approximations are both constant. Given that they are approximations of the same function, then we need to require that the inner and outer expansions are equal in this region. In more mathematical terms, the requirement we will impose on these two expansions is

$$\lim_{\bar{x} \rightarrow \infty} Y_0 = \lim_{x \rightarrow 0} y_0. \quad (2.64)$$

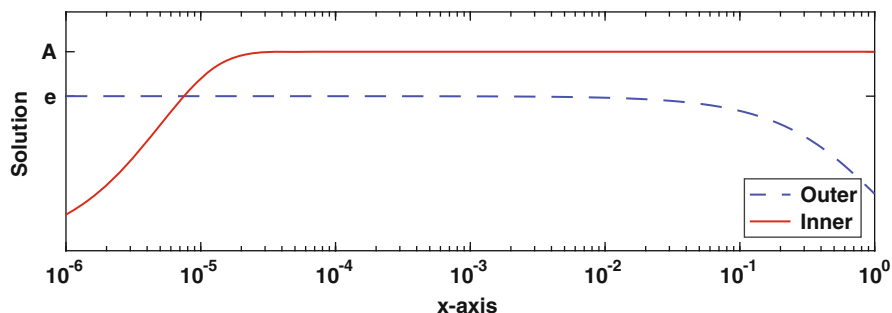
This is called the *matching condition*, and it is often written in the form

$$Y_0(\infty) = y_0(0). \quad (2.65)$$

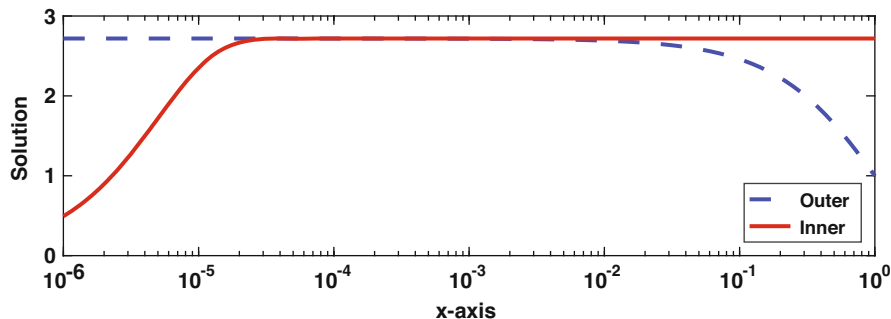
With this we conclude  $A = e$  and the resulting functions are plotted in Fig. 2.11 for  $\varepsilon = 10^{-4}$ . The overlap domain is clearly seen in this figure.

### Step 4: Composite Approximation

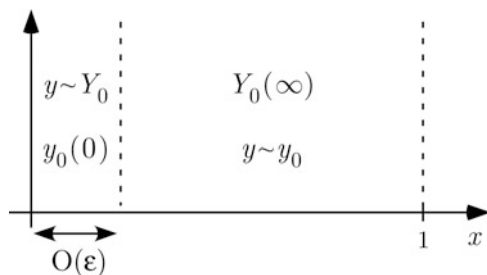
The approximation of the solution we have comes in two pieces, one that applies near  $x = 0$  and another that works everywhere else. Because neither can be used over the entire interval we say that they are not uniformly valid for  $0 \leq x \leq 1$ . The question we consider now is whether we can combine them in some way to



**Fig. 2.10** Graph of the inner approximation (2.63), and the outer approximation (2.56), before matching



**Fig. 2.11** Graph of the inner approximation (2.63), and the outer approximation (2.56), after matching in the particular case of when  $\varepsilon = 10^{-4}$ . Note the overlap region where the two approximations produce, approximately, the same result

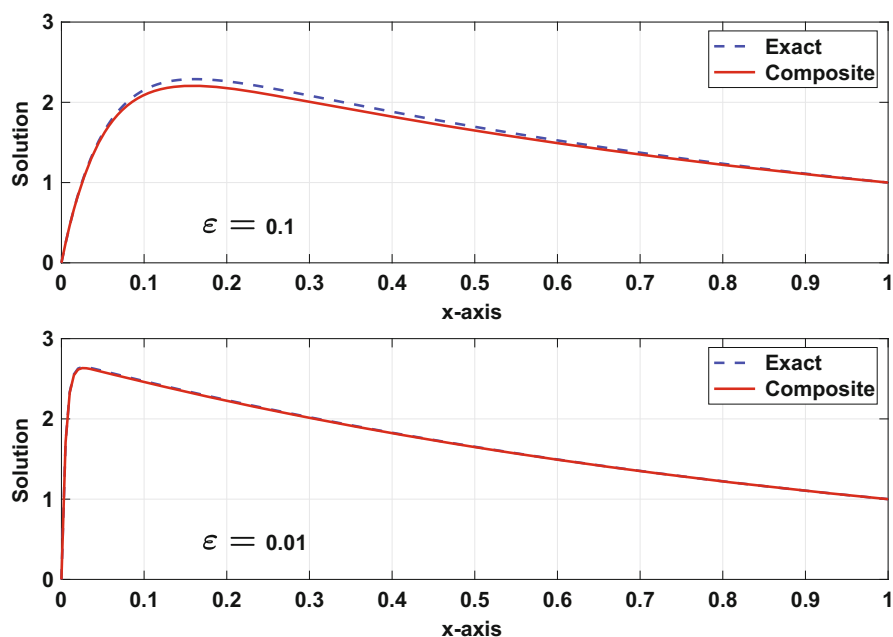


**Fig. 2.12** Sketch of the boundary layer and outer region, and the values of the approximations in those regions

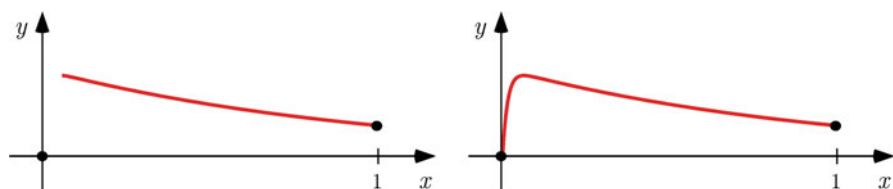
produce a uniform approximation, that is, one that works over the entire interval. The position we are in is summarized in Fig. 2.12. The inner and outer solutions are constant outside the region where they are used to approximate the solution, and the constant is the same for both solutions. The value of the constant can be written as either  $y_0(0)$  or as  $Y_0(\infty)$ , and the fact that they are equal is a consequence of the matching condition (2.64). This observation can be used to construct a uniform approximation. Namely, we just add the approximations together and then subtract the constant. The result is

$$\begin{aligned} y &\sim y_0(x) + Y_0(\bar{x}) - y_0(0) \\ &= e^{1-x} - e^{1-2x/\varepsilon}. \end{aligned} \quad (2.66)$$

This is known as a *composite approximation* and it is valid for  $0 \leq x \leq 1$ . To demonstrate its effectiveness it is plotted in Fig. 2.13 along with the exact solution for  $\varepsilon = 10^{-1}$  and for  $\varepsilon = 10^{-2}$ . It is evident from this figure that we have constructed a relatively simple expression that is a very good approximation of the solution over the entire interval.



**Fig. 2.13** Graph of the exact solution (2.52) and composite approximation (2.66) for two values of  $\varepsilon$



**Fig. 2.14** To sketch of the solution of (2.49)–(2.51) one first sketches the outer solution (left plot), and then draws in the steep, monotonic, boundary layer portion (right plot)

### Sketching the Solution

The way the composite approximation is determined can also be used to sketch the solution fairly easily. You first sketch the outer solution, outside of the boundary layer region, as illustrated on the left in Fig. 2.14. Now, boundary layer solutions are usually strictly monotonic functions that are very steep, and they connect the outer solution with the boundary condition associated with the boundary layer. So, you simply draw in such a curve, making sure to make a smooth transition between the outer solution and the boundary layer portion. The result is shown on the right in Fig. 2.14.

It is possible to produce the sketch of the solution shown on the right in Fig. 2.14 if you know where the boundary is located and you have determined the outer

solution. For problems where you do not know where the layer is, this sort of sketch can be very useful to help determine the layer's location. How this is done is explained in Holmes (2013b).

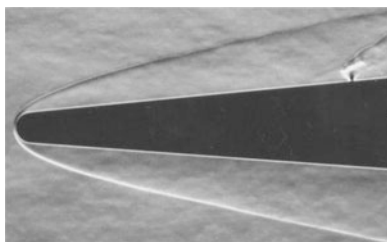
### 2.5.1 Endnotes

One of the characteristics of a boundary layer is that its width goes to zero as  $\varepsilon \rightarrow 0$ , yet the change in the solution across the layer does not go to zero. This type of behavior occurs in a wide variety of problems, although the terminology changes depending on the application and particular type of problem. For example, there are problems where the jump occurs in the interval  $0 < x < 1$ , a situation known as an interior layer. They are also not limited to BVPs and arise in IVPs, PDEs, etc. An example of this is shown in Fig. 2.15. The boundary layer is the thin white region on the surface of the object. In this layer the air velocity changes rapidly, from zero on the object to the large value in the outer flow. The parabolic curve that appears to be attached to the front of the object is a shock wave. The pressure undergoes a rapid change across the shock, and for this reason it is an example of an interior layer.

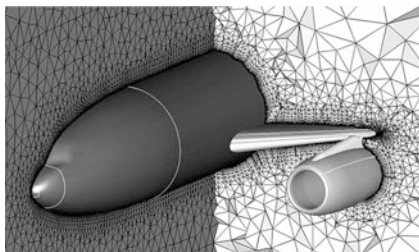
The presence of a boundary layer is an issue when finding the numerical solution. As an example, Fig. 2.16 shows the grid system used to solve the equations for the air flow over an object, in this case an airplane. The presence of a boundary layer necessitates the use of a large number of grid points near the surface, which greatly adds to the computational effort needed to solve the problem.

Another important comment to make concerns the existence of a boundary layer in the solution. In particular, an  $\varepsilon$  multiplying the highest derivative is not

**Fig. 2.15** Image of high speed flow, from left to right, over a fixed, wedge-shaped object. The thin white region on the surface of the object is the boundary layer. The parabolic curve is a shock wave, a topic which is studied in Chap. 5



**Fig. 2.16** Grid refinement needed near the boundary to numerically calculate the air flow over an airplane (Steinbrenner and Abelanet 2007)



a guarantee of a boundary, or interior, layer. A simple example is  $\varepsilon y'' + y = 0$ , for which the general solution is  $y = a \sin(x/\sqrt{\varepsilon}) + b \cos(x/\sqrt{\varepsilon})$ . In this case, instead of containing a rapidly decaying exponential function that is characteristic of a boundary layer, the solution consists of rapidly varying oscillatory functions. The approximation method most often used in such situations is known as the WKB method. We will only scratch the surface of this subject, and a more extensive study of this can be found in Holmes (2013b).

Finally, as stated earlier, it is typical that in physical problems that you know where the boundary layer is located. Figures 2.15 and 2.16 are illustrations of this observation.

## 2.6 Examples Involving Boundary Layers

Three examples are considered. The steps mimic those in the previous example, and so the presentation here is briefer. It is important to remember in the analysis to follow that the goal is to find a first term approximation of the solution over the interval  $0 \leq x \leq 1$ .

### 2.6.1 Example 1: Layer at Left End

Consider the problem of solving

$$\varepsilon^2 y'' + y' + \varepsilon y^2 = 1 + x, \quad \text{for } 0 < x < 1, \quad (2.67)$$

where the boundary conditions are

$$y(0) = -1, \quad (2.68)$$

and

$$y(1) = 2. \quad (2.69)$$

The layer in this problem is at  $x = 0$ , and the objective is to drive an approximation of the solution for  $0 \leq x \leq 1$ .

The purpose of this example is to demonstrate how several of the steps in the derivation of the composite expansion can be simplified. If questions arise as to why, or how, the steps occur, you should review the similar, but more expansive, derivation in Sect. 2.5.

### Step 1: Outer Solution

Assuming that  $y \sim y_0(x) + \varepsilon y_1(x) + \dots$ , then the differential equation for  $y_0$  can be found by simply setting  $\varepsilon = 0$  in (2.67). Also, given that the layer is at  $x = 0$ , then  $y_0$  must satisfy (2.69). The resulting problem to solve is:

$$O(1) \quad y_0' = 1 + x$$

$$y_0(1) = 2$$

The general solution of the differential equation is  $y_0 = x + \frac{1}{2}x^2 + a$ , where  $a$  is an arbitrary constant. Since  $y_0(1) = 2$  it follows that  $a = 1/2$ , and so  $y_0(x) = x + \frac{1}{2}(x^2 + 1)$ .

### Step 2: Boundary Layer Solution

The boundary layer coordinate is  $\bar{x} = x/\varepsilon^\gamma$ , where  $\gamma > 0$ . Substituting this into (2.67), and letting  $Y(\bar{x})$  denote the solution when using  $\bar{x}$  as the independent variable, we get that

$$\varepsilon^{2-2\gamma} Y'' + \varepsilon^{-\gamma} Y' + \varepsilon Y^2 = 1 + \varepsilon^\gamma \bar{x}. \quad (2.70)$$

$$\textcircled{1} \qquad \qquad \textcircled{2} \qquad \qquad \textcircled{3} \quad \textcircled{4} \quad \textcircled{5}$$

Because of the  $-\gamma$  exponent in  $\textcircled{2}$ , it follows immediately that  $\textcircled{3} \ll \textcircled{2}$ ,  $\textcircled{4} \ll \textcircled{2}$ , and  $\textcircled{5} \ll \textcircled{2}$ . In other words, the only possible balance for the boundary layer involves  $\textcircled{1}$  and  $\textcircled{2}$ . From this, we get that  $\gamma = 2$ , and so the problem becomes

$$Y'' + Y' + \varepsilon^3 Y^2 = \varepsilon + \varepsilon^2 \bar{x},$$

where  $Y(0) = -1$ . Assuming  $Y(\bar{x}) \sim Y_0(\bar{x}) + \varepsilon Y_1(\bar{x}) + \dots$ , we obtain the following problem:

$$O(1) \quad Y_0'' + Y_0' = 0$$

$$Y_0(0) = -1$$

The general solution of the differential equation is  $Y_0 = A + B e^{-\bar{x}}$ , where  $A$  and  $B$  are arbitrary constants. With the given boundary condition this reduces to  $Y_0 = -1 - B + B e^{-\bar{x}}$ .

### Step 3: Matching

The requirement is that  $Y_0(\infty) = y_0(0)$ , which means  $-1 - B = \frac{1}{2}$ . So,  $B = -\frac{3}{2}$ , and  $Y_0 = \frac{1}{2} - \frac{3}{2} e^{-\bar{x}}$ .

### Step 4: Composite Expansion

The resulting composite expansion is

$$\begin{aligned} y &\sim y_0(x) + Y_0(\bar{x}) - y_0(0) \\ &= x + \frac{1}{2}(x^2 + 1) - \frac{3}{2} e^{-\bar{x}}. \end{aligned}$$



### 2.6.2 Example 2: Layer at Right End

Consider the problem of solving

$$\varepsilon y'' - y' + xy = 0, \quad \text{for } 0 < x < 1, \quad (2.71)$$

where the boundary conditions are

$$y(0) = 3, \quad (2.72)$$

and

$$y(1) = -1. \quad (2.73)$$

The layer in this problem is at  $x = 1$ , so we will require the outer approximation to satisfy the boundary condition at  $x = 0$ .

#### Step 1: Outer Solution

Assuming that  $y \sim y_0(x) + \varepsilon y_1(x) + \dots$ , then the differential equation for  $y_0$  can be found by simply setting  $\varepsilon = 0$  in (2.71). Also, given that the layer is at  $x = 1$ , then  $y_0$  must satisfy (2.72). The resulting problem to solve is:

$$\begin{aligned} O(1) \quad & -y_0' + xy_0 = 0 \\ & y_0(0) = 3 \end{aligned}$$

The general solution of the differential equation is  $y_0 = ae^{x^2/2}$ , where  $a$  is an arbitrary constant. Since  $y_0(0) = 3$  it follows that  $a = 3$  and  $y_0(x) = 3e^{x^2/2}$ .

#### Step 2: Boundary Layer Solution

Since the layer is at  $x = 1$ , the boundary layer coordinate is

$$\tilde{x} = \frac{x - 1}{\varepsilon^\gamma}. \quad (2.74)$$

To explain, the boundary layer coordinate must be centered at the endpoint where the boundary layer is located, hence the numerator  $x - 1$ . It is also assumed to have a width that is  $O(\varepsilon^\gamma)$ , which explains the denominator. One consequence of this is that  $x = 1$  now corresponds to  $\tilde{x} = 0$ , and  $x < 1$  means that  $\tilde{x} < 0$ . As for the transformation of the derivatives, using the chain rule,

$$\frac{d}{dx} = \frac{1}{\varepsilon^\gamma} \frac{d}{d\tilde{x}} \quad \text{and} \quad \frac{d^2}{dx^2} = \frac{1}{\varepsilon^{2\gamma}} \frac{d^2}{d\tilde{x}^2}.$$

Lastly, note that it is possible to write (2.74) as

$$x = 1 + \varepsilon^\gamma \tilde{x}. \quad (2.75)$$

We will designate the solution as  $\tilde{Y}(\tilde{x})$  when using  $\tilde{x}$  as the independent variable. With this, the differential equation becomes

$$\underbrace{\varepsilon^{1-2\gamma}}_{\textcircled{1}} \underbrace{\tilde{Y}''}_{\textcircled{2}} - \underbrace{\varepsilon^{-\gamma}}_{\textcircled{3}} \underbrace{\tilde{Y}'}_{\textcircled{4}} + (1 + \tilde{x}\varepsilon^\gamma)\tilde{Y} = 0. \quad (2.76)$$

As in the last example, because of the  $-\gamma$  exponent in  $\textcircled{2}$ , it follows immediately that  $\textcircled{3} \ll \textcircled{2}$  and  $\textcircled{4} \ll \textcircled{2}$ . In other words, the only possible balance for the boundary layer involves  $\textcircled{1}$  and  $\textcircled{2}$ . From this, we get that  $\gamma = 1$ , and so the problem becomes

$$\tilde{Y}'' - \tilde{Y}' + \varepsilon(1 + \tilde{x}\varepsilon^\gamma)\tilde{Y} = 0. \quad (2.77)$$

Also, the boundary condition  $y(1) = -1$  now becomes  $\tilde{Y}(0) = -1$ .

Assuming  $\tilde{Y}(\tilde{x}) \sim \tilde{Y}_0(\tilde{x}) + \varepsilon\tilde{Y}_1(\tilde{x}) + \dots$ , we obtain the following problem.

$$\begin{aligned} O(1) \quad & \tilde{Y}_0'' - \tilde{Y}_0' = 0 \\ & \tilde{Y}_0(0) = -1 \end{aligned}$$

The general solution of the differential equation is  $\tilde{Y}_0 = A + Be^{\tilde{x}}$ , where  $A$  and  $B$  are arbitrary constants. With the given boundary condition this reduces to  $\tilde{Y}_0 = -1 + B(e^{\tilde{x}} - 1)$ .

### Step 3: Matching

According to the matching requirement, the boundary layer solution going into the outer region should give the same value as the outer solution going into the boundary layer. Mathematically this means that

$$\lim_{\tilde{x} \rightarrow -\infty} \tilde{Y}_0 = \lim_{x \rightarrow 1} y_0, \quad (2.78)$$

or equivalently  $\tilde{Y}_0(-\infty) = y_0(1)$ . Since  $\tilde{Y}_0(-\infty) = -1 - B$  and  $y_0(1) = 3e^{1/2}$ , it follows that  $B = -1 - 3e^{1/2}$ .

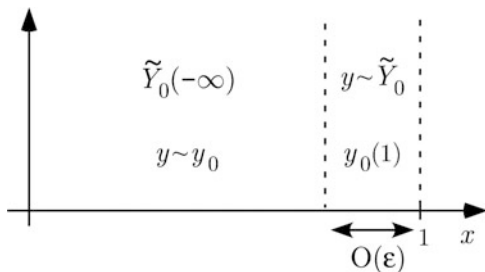
### Step 4: Composite Expansion

A sketch of where our approximations are valid is shown in Fig. 2.17. Using the same reasoning that lead to (2.66), we have that a composite expansion is

$$\begin{aligned} y & \sim y_0(x) + \tilde{Y}_0(\tilde{x}) - y_0(1) \\ & = 3e^{x^2/2} - (1 + 3e^{1/2})e^{(x-1)/\varepsilon}. \end{aligned} \quad (2.79)$$

This is the sought first term approximation of the solution that holds for  $0 \leq x \leq 1$ .

**Fig. 2.17** Sketch of the boundary layer and outer regions and the values of the approximations in those regions



### 2.6.3 Example 3: Boundary Layer at Both Ends

As a third boundary layer example we will consider the boundary value problem

$$\varepsilon^2 y'' + \varepsilon x y' - y = 1 - 2x, \quad \text{for } 0 < x < 1, \quad (2.80)$$

where the boundary conditions are

$$y(0) = 2, \quad (2.81)$$

and

$$y(1) = -1. \quad (2.82)$$

As will be evident almost immediately, this problem has a boundary layer at each end of the interval.

#### Step 1: Outer Solution

Assuming  $y \sim y_0(x) + \varepsilon y_1(x) + \dots$  one finds from the differential equation that  $y_0 = 2x - 1$ . This function is incapable of satisfying either boundary condition, hence the reason for having boundary layers at both ends.

#### Steps 2 and 3: Boundary Layer Solutions and Matching

Given that there is a layer at each end we need to split this step into two parts.

**(a) Layer at  $x = 0$**  In this region we will denote the solution as  $Y(\bar{x})$ . The boundary coordinate is the same as before. Setting  $\bar{x} = x/\varepsilon^\gamma$  and using the formulas in (2.58) and (2.59) the differential equation (2.80) becomes

$$\varepsilon^{2-2\gamma} Y'' + \varepsilon \bar{x} Y' - Y = 1 - 2\varepsilon^\gamma \bar{x}. \quad (2.83)$$

①                      ②                      ③                      ④                      ⑤

First note that terms ③ and ④ are the same order. Also, ②  $\ll$  ③ and ⑤  $\ll$  ③. Consequently, the only possible balance for the boundary layer involves ①, ③, and ④. From this, we get that  $\gamma = 1$ , and so the problem becomes

$$Y'' + \varepsilon \bar{x} Y' - Y = 1 - 2\varepsilon \bar{x}, \quad (2.84)$$

where  $Y(0) = 2$ . Assuming  $Y(\bar{x}) \sim Y_0(\bar{x}) + \varepsilon Y_1(\bar{x}) + \dots$  we obtain the following problem to solve:

$$O(1) \quad \begin{aligned} Y_0'' - Y_0 &= 1 \\ Y_0(0) &= 2 \end{aligned}$$

The general solution of the differential equation is  $Y_0 = -1 + Ae^{\bar{x}} + Be^{-\bar{x}}$ , where  $A$  and  $B$  are arbitrary constants. With the given boundary condition this reduces to  $Y_0 = -1 + Ae^{\bar{x}} + (3 - A)e^{-\bar{x}}$ .

The boundary layer solution must match with the outer solution. The requirement is  $Y_0(\infty) = y_0(0)$ . Since  $\lim_{\bar{x} \rightarrow \infty} e^{\bar{x}} = \infty$ , then the matching condition requires that  $-1 + A \cdot \infty + (3 - A) \cdot 0 = -1$ . This holds if  $A = 0$ , and so the first term approximation in this boundary layer is

$$Y_0(\bar{x}) = -1 + 3e^{-\bar{x}}. \quad (2.85)$$

**(b) Layer at  $x = 1$**  In this region we will denote the solution as  $\tilde{Y}(\tilde{x})$ . The boundary layer in this case is located at  $x = 1$ , and so the coordinate will be centered at this point. In particular, we let

$$\tilde{x} = \frac{x - 1}{\varepsilon^\gamma}. \quad (2.86)$$

The differentiation formulas are similar to those in (2.58) and (2.59). Also, we have that  $x = 1 + \varepsilon^\gamma \tilde{x}$ . With this the differential equation (2.80) becomes

$$\begin{aligned} \varepsilon^{2-2\gamma} \tilde{Y}'' + \varepsilon^{1-\gamma} (1 + \varepsilon^\gamma \tilde{x}) \tilde{Y}' - \tilde{Y} &= -1 - 2\varepsilon^\gamma \tilde{x}. \end{aligned} \quad (2.87)$$

①
②
③
④

Note that the two unlabeled  $\varepsilon^\gamma$  terms are not considered because they are higher order as compared to ② and ④, respectively. As for balancing, if ①  $\sim$  ②, then  $\gamma = 1$ . In this case, all four terms ①, ②, ③, and ④ are  $O(1)$ . Consequently, with  $\gamma = 1$ , the differential equation becomes

$$\tilde{Y}'' + (1 + \varepsilon \tilde{x}) \tilde{Y}' - \tilde{Y} = -1 - 2\varepsilon \tilde{x}, \quad (2.88)$$

and the boundary condition to satisfy is

$$\tilde{Y}(0) = -1. \quad (2.89)$$

As a reminder,  $\tilde{Y}$  is evaluated at  $\tilde{x} = 0$  because  $x = 1$  corresponds to  $\tilde{x} = 0$ .

Assuming  $\tilde{Y}(\tilde{x}) \sim \tilde{Y}_0(\tilde{x}) + \varepsilon \tilde{Y}_1(\tilde{x}) + \dots$  we obtain the following problem to solve.

$$O(1) \quad \tilde{Y}_0'' + \tilde{Y}_0' - \tilde{Y}_0 = -1$$

$$\tilde{Y}_0(0) = -1$$

The general solution of the differential equation is  $\tilde{Y}_0 = 1 + Ae^{r_+\tilde{x}} + Be^{r_-\tilde{x}}$ , where  $r_{\pm} = (-1 \pm \sqrt{5})/2$  and  $A, B$  are arbitrary constants. With the given boundary condition this reduces to  $\tilde{Y}_0 = 1 + Ae^{r_+\tilde{x}} - (2+A)e^{r_-\tilde{x}}$ .

This boundary layer solution must match with the outer solution. The requirement is  $\tilde{Y}_0(-\infty) = y_0(1)$ . Given that  $r_+ > 0$  and  $r_- < 0$ , then  $\lim_{\tilde{x} \rightarrow -\infty} e^{r_+\tilde{x}} = \infty$  and  $\lim_{\tilde{x} \rightarrow -\infty} e^{r_-\tilde{x}} = 0$ . For  $\tilde{Y}_0$  to be able to match with the outer solution we must set  $2 + A = 0$ . With this our first term approximation in this boundary layer is

$$\tilde{Y}_0(\tilde{x}) = 1 - 2e^{r_+\tilde{x}}. \quad (2.90)$$

#### Step 4: Composite

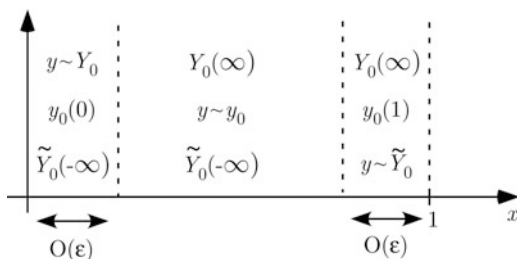
As in the previous examples, it is possible to combine the three approximations we have derived to produce a uniform approximation. The situation is shown schematically in Fig. 2.18. It is seen that in each region the two approximations not associated with that region add to  $y_0(0) + y_0(1)$ . This means we simply add the three approximations together and subtract  $y_0(0) + y_0(1)$ . In other words,

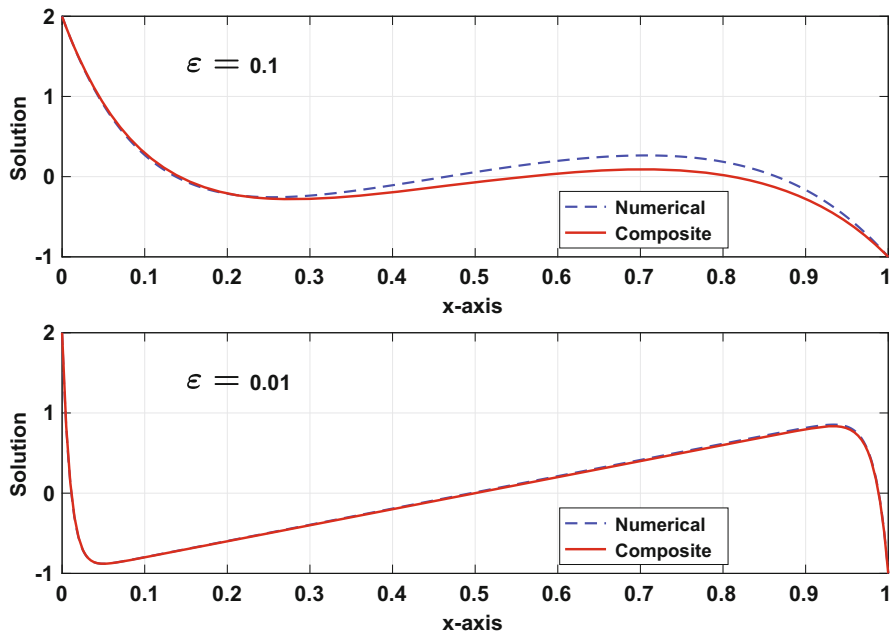
$$y \sim y_0(x) + Y_0(\bar{x}) + \tilde{Y}_0(\tilde{x}) - y_0(0) - y_0(1)$$

$$= -1 + 2x + 3e^{-x/\varepsilon} - 2e^{r_+(x-1)/\varepsilon}. \quad (2.91)$$

This function is a composite expansion of the solution and it is valid for  $0 \leq x \leq 1$ . To demonstrate its effectiveness, the composite approximation is plotted in Fig. 2.19 along with the numerical solution for  $\varepsilon = 10^{-1}$  and for  $\varepsilon = 10^{-2}$ . As is evident, the approximation is not bad for  $\varepsilon = 10^{-1}$ , and it is quite good  $\varepsilon = 10^{-2}$ . It is also expected that the approximation improves for smaller values of  $\varepsilon$ .

**Fig. 2.18** Sketch of the three regions and the values of the approximations in those regions. Note that  $y_0(0) = Y_0(\infty) = -1$  and  $y_0(1) = \tilde{Y}_0(-\infty) = 1$





**Fig. 2.19** Graph of the numerical solution of the boundary value problem (2.80)–(2.82) and the composite approximation of the solution (2.91)

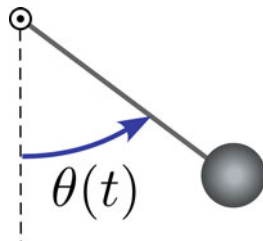
## 2.7 Multiple Scales

As the last three examples have demonstrated, the presence of a boundary layer limits the region over which an approximation can be used. Said another way, the inner and outer approximations are not uniformly valid over the entire interval. The tell-tale sign that this is going to happen is that when  $\varepsilon = 0$  the highest derivative in the problem is lost. However, the lack of uniformity can occur in other ways and one investigated here relates to changes in the solution as a function of time. It is easier to explain what happens by working out a typical example. For this we use the pendulum problem. Letting  $\theta(t)$  be the angular deflection made by the pendulum, as shown in Fig. 2.20, the problem is

$$\theta'' + \sin \theta = 0, \quad (2.92)$$

where

$$\theta(0) = \varepsilon, \quad (2.93)$$

**Fig. 2.20** Pendulum example

and

$$\theta'(0) = 0. \quad (2.94)$$

The equation of motion (2.92) comes from Newton's second law,  $F = ma$ , where the external forcing  $F$  is gravity. It is assumed the initial angle is small, and this is the reason for the initial condition (2.93). It is also assumed that the pendulum starts from rest, so the initial velocity (2.94) is zero.

Although the problem is difficult to solve we have at least some idea of what the solution looks like because of everyday experience with a pendulum (e.g., watching a grandfather clock or using a swing). Starting out with the given initial conditions, the pendulum should simply oscillate back and forth. A real pendulum will eventually stop due to damping, but we have not included this in the model so our pendulum should go forever.

### 2.7.1 Regular Expansion

The fact that the small parameter is in the initial condition, and not in the differential equation, is a bit different from what we had in the previous section, but we are still able to use our usual approximation methods. The appropriate expansion in this case is

$$\theta(t) \sim \varepsilon(\theta_0(t) + \varepsilon^\alpha \theta_1(t) + \dots). \quad (2.95)$$

The  $\varepsilon$  multiplying the series is there because of the initial condition. If we did not have it, and tried  $\theta = \theta_0 + \varepsilon^\alpha \theta_1 + \dots$ , we would find that  $\theta_0 = 0$  and  $\alpha = 1$ . The assumption in (2.95) is made simply to avoid all the work to find that the first term in the expansion is just zero. Before substituting (2.95) into the problem recall  $\sin(x) = x - \frac{1}{6}x^3 + \dots$  when  $x$  is close to zero. This means, because the  $\theta$  in (2.95) is close to zero,

$$\begin{aligned}
\sin \theta &\sim \sin(\varepsilon(\theta_0 + \varepsilon^\alpha \theta_1 + \dots)) \\
&\sim (\varepsilon\theta_0 + \varepsilon^{\alpha+1}\theta_1 + \dots) - \frac{1}{6}(\varepsilon\theta_0 + \dots)^3 + \dots \\
&\sim \varepsilon\theta_0 + \varepsilon^{\alpha+1}\theta_1 - \frac{1}{6}\varepsilon^3\theta_0^3 + \dots.
\end{aligned} \tag{2.96}$$

With this the equation of motion (2.92) becomes

$$\varepsilon\theta_0'' + \varepsilon^{\alpha+1}\theta_1'' + \dots + \varepsilon\theta_0 + \varepsilon^{\alpha+1}\theta_1 - \frac{1}{6}\varepsilon^3\theta_0^3 + \dots = 0, \tag{2.97}$$

and the initial conditions are

$$\varepsilon\theta_0(0) + \varepsilon^{\alpha+1}\theta_1(0) + \dots = \varepsilon, \tag{2.98}$$

and

$$\varepsilon\theta_0'(0) + \varepsilon^{\alpha+1}\theta_1'(0) + \dots = 0. \tag{2.99}$$

Proceeding in the usual manner yields the following problem.

$$\begin{aligned}
O(\varepsilon) \quad &\theta_0'' + \theta_0 = 0 \\
&\theta_0(0) = 1, \theta_0'(0) = 0
\end{aligned}$$

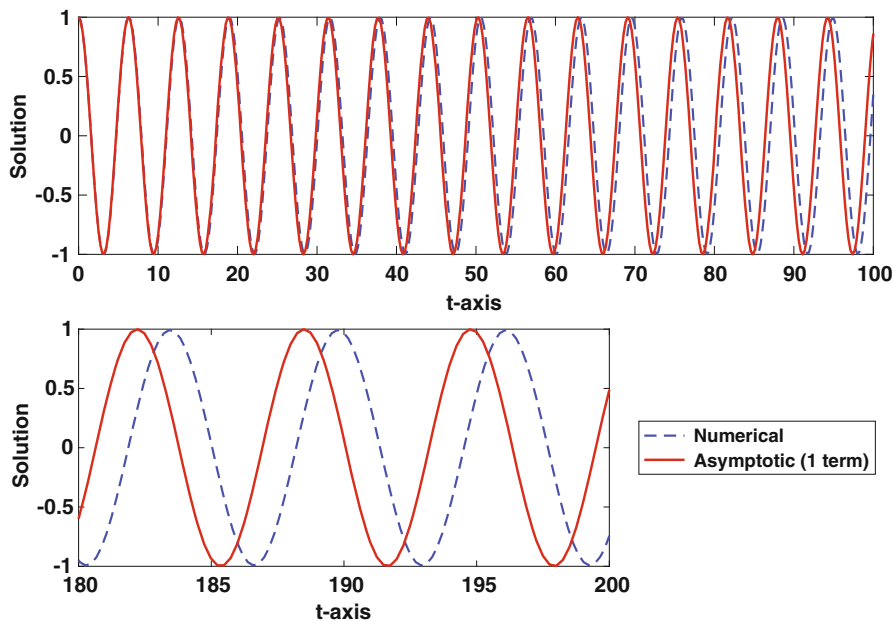
The general solution of the differential equation is  $\theta_0 = a \cos(t) + b \sin(t)$ , where  $a, b$  are arbitrary constants. It is possible to write this solution in the more compact form  $\theta_0 = A \cos(t+B)$ , where  $A, B$  are arbitrary constants. As will be explained later, there is a reason for why the latter form is preferred in this problem. With this, and the initial conditions, it is found that  $\theta_0 = \cos(t)$ .

The plot of the one-term approximation,  $\theta \sim \varepsilon \cos(t)$ , and the numerical solution are shown in Fig. 2.21. The asymptotic approximation describes the solution accurately at the start, and reproduces the amplitude very well over the entire time interval. What it has trouble with is matching the phase and this is evident in the lower plot in Fig. 2.21. One additional comment to make is the value for  $\varepsilon$  used in Fig. 2.21 is not particularly small, so getting an approximation that is not very accurate is no surprise. However, if a smaller value is used the same difficulty arises. The difference is that the first term approximation works over a longer time interval but eventually the phase difference seen in Fig. 2.21 occurs.

In looking to correct the approximation to reduce the phase error we calculate the second term in the expansion. With the given  $\theta_0$  there is an  $\varepsilon^3\theta_0^3$  term in (2.97). To balance this we use the  $\theta_1$  term in the expansion and this requires  $\alpha = 2$ . With this we have the following problem to solve.

$$\begin{aligned}
O(\varepsilon^3) \quad &\theta_1'' + \theta_1 = \frac{1}{6}\theta_0^3 \\
&\theta_1(0) = 0, \theta_1'(0) = 0
\end{aligned}$$





**Fig. 2.21** Graph of the numerical solution of the pendulum problem (2.80)–(2.82) and the first term in the regular perturbation approximation (2.95). Shown are the solutions for  $0 \leq t \leq 100$ , as well as a close up of the solutions near  $t = 200$ . In the calculation  $\varepsilon = \frac{1}{3}$  and both solutions have been divided by  $\varepsilon = \frac{1}{3}$

The method of undetermined coefficients can be used to find a particular solution of this equation. This requires the identity  $\cos^3(t) = \frac{1}{4}(3 \cos(t) + \cos(3t))$ , in which case the differential equation becomes

$$\theta_1'' + \theta_1 = \frac{1}{24}(3 \cos(t) + \cos(3t)). \quad (2.100)$$

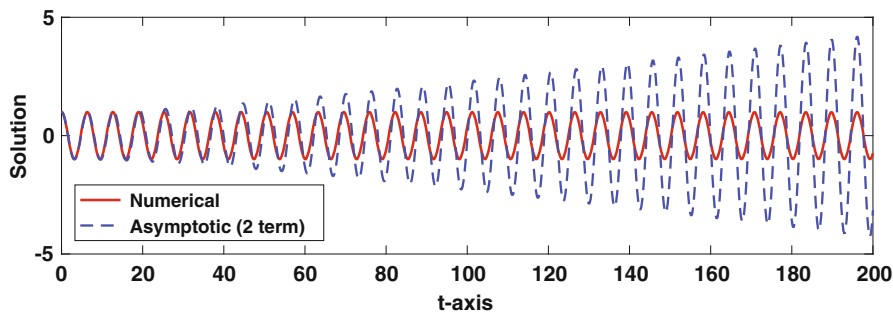
With this the general solution is found to be

$$\theta_1 = a \cos(t) + b \sin(t) + \frac{1}{16}t \sin(t) - \frac{1}{192} \cos(3t),$$

where  $a, b$  are arbitrary constants. From the initial conditions this reduces to  $\theta_1 = \frac{1}{192} \cos(t) - \frac{1}{192} \cos(3t) + \frac{1}{16}t \sin(t)$ .

The plot of the two term approximation,

$$\theta \sim \varepsilon \cos(t) + \varepsilon^3 \left[ \frac{1}{192} \cos(t) - \frac{1}{192} \cos(3t) + \frac{1}{16}t \sin(t) \right], \quad (2.101)$$



**Fig. 2.22** Graph of the numerical solution of the pendulum problem (2.80)–(2.82) and the regular perturbation approximation (2.91). In the calculation  $\varepsilon = \frac{1}{3}$  and the solution has been divided by  $\varepsilon = \frac{1}{3}$

and the numerical solution are shown in Fig. 2.22. It is clear from this that we have been less than successful in improving the approximation. The culprit here is the  $t \sin(t)$  term in (2.101). As time increases its contribution grows, and it eventually gets as large as the first term in the expansion. Because of this it is called a *secular term*, and it causes the expansion not to be uniformly valid for  $0 \leq t < \infty$ . This problem would not occur if time were limited to a finite interval, as happened in the projectile problem. However, for the pendulum there is no limit on time and this means the expansion is restricted to when it can be used.

One last comment to make concerns how this term ended up in the expansion in the first place. In the differential equation for  $\theta_1$ , given in (2.100), the right-hand side contains  $\cos(t)$  and this is a solution of the associated homogeneous equation. It is this term that produces the  $t \sin(t)$  in the expansion and it is this term we would like to prevent from appearing in the problem.

### 2.7.2 Multiple Scales Expansion

What is happening is that there is a slow change in the solution that the first term approximation is unable to describe. In effect there are two time scales acting in this problem. One is the basic period of oscillation, as seen in Fig. 2.21, and the other is a slow time scale over which the phase changes. Our approximation method will be based on this observation. We will explicitly assume there are two concurrent time scales, given as

$$t_1 = t, \quad (2.102)$$

$$t_2 = \varepsilon^\gamma t. \quad (2.103)$$

The value of  $\gamma$  is not known yet, and we will let the problem tell us the value as we construct the approximation. Based on this assumption it is not surprising that the method is called two-timing, or the method of multiple scales.

To illustrate the idea underlying two-timing, consider the function

$$u = e^{-3\epsilon t} \sin(5t).$$

This consists of an oscillatory function, with a slowly decaying amplitude. This can be written using the two-timing variables as

$$u = e^{-3t_2} \sin(5t_1),$$

where  $\gamma = 1$ .

The change of variables in (2.102) and (2.103) is reminiscent of the boundary layer problems in the previous section. The difference here is that we are not separating the time axis into separate regions but, rather, using two time scales together. As we will see, this has a profound effect on how we construct the approximation.

To determine how the change of variables affects the time derivative, we have, using the chain rule,

$$\begin{aligned} \frac{d}{dt} &= \frac{dt_1}{dt} \frac{\partial}{\partial t_1} + \frac{dt_2}{dt} \frac{\partial}{\partial t_2} \\ &= \frac{\partial}{\partial t_1} + \epsilon^\gamma \frac{\partial}{\partial t_2}. \end{aligned} \quad (2.104)$$

The second derivative is

$$\begin{aligned} \frac{d^2}{dt^2} &= \left( \frac{\partial}{\partial t_1} + \epsilon^\gamma \frac{\partial}{\partial t_2} \right) \left( \frac{\partial}{\partial t_1} + \epsilon^\gamma \frac{\partial}{\partial t_2} \right) \\ &= \frac{\partial^2}{\partial t_1^2} + 2\epsilon^\gamma \frac{\partial^2}{\partial t_1 \partial t_2} + \epsilon^{2\gamma} \frac{\partial^2}{\partial t_2^2}. \end{aligned} \quad (2.105)$$

The steps used to construct an asymptotic approximation of the solution will closely follow what we did earlier. It should be kept in mind during the derivation that the sole reason for introducing  $t_2$  is to prevent a secular term from appearing in the second term.

With the introduction of a second time variable, the expansion is assumed to have the form

$$\theta \sim \epsilon(\theta_0(t_1, t_2) + \epsilon^\alpha \theta_1(t_1, t_2) + \cdots). \quad (2.106)$$

The only difference between this and the regular expansion (2.95) used earlier is that the terms are allowed to depend on both time variables. When this is substituted

into the differential equation we obtain an expression similar to (2.97), except the time derivatives are given in (2.104) and (2.105). Specifically, we get

$$\varepsilon \frac{\partial^2}{\partial t_1^2} \theta_0 + \varepsilon^{\alpha+1} \frac{\partial^2}{\partial t_1^2} \theta_1 + 2\varepsilon^{\gamma+1} \frac{\partial^2}{\partial t_1 \partial t_2} \theta_0 + \cdots + \varepsilon \theta_0 + \varepsilon^{\alpha+1} \theta_1 - \frac{1}{6} \varepsilon^3 \theta_0^3 + \cdots = 0, \quad (2.107)$$

and the initial conditions are

$$\varepsilon \theta_0(0, 0) + \varepsilon^{\alpha+1} \theta_1(0, 0) + \cdots = \varepsilon, \quad (2.108)$$

and

$$\varepsilon \frac{\partial}{\partial t_1} \theta_0(0, 0) + \varepsilon^{\alpha+1} \frac{\partial}{\partial t_1} \theta_1(0, 0) + \varepsilon^{\gamma+1} \frac{\partial}{\partial t_2} \theta_0(0, 0) + \cdots = 0. \quad (2.109)$$

Proceeding in the usual manner yields the following problem.

$$O(\varepsilon) \quad \frac{\partial^2}{\partial t_1^2} \theta_0 + \theta_0 = 0$$

$$\theta_0(0, 0) = 1, \quad \frac{\partial}{\partial t_1} \theta_0(0, 0) = 0$$

The general solution of the differential equation is  $\theta_0 = A(t_2) \cos(t_1 + B(t_2))$ , where  $A, B$  are arbitrary functions of  $t_2$ . The effects of the second time variable are seen in this solution, because the coefficients are now functions of the second time variable. To satisfy the initial conditions we need  $A(0) \cos(B(0)) = 1$  and  $A(0) \sin(B(0)) = 0$ . From this we have that  $A(0) = 1$  and  $B(0) = 0$ .

In the differential equation (2.107), with the  $O(\varepsilon)$  terms out of the way, the next term to consider is  $\varepsilon^3 \theta_0^3$ . The only terms we have available to balance with this have order  $\varepsilon^{\alpha+1}$  and  $\varepsilon^{\gamma+1}$ . To determine which terms to use we can use a balance argument, similar to what was done for boundary layers. It is found that both terms are needed and this means  $\alpha + 1 = \gamma + 1 = 3$ . This is an example of a distinguished balance. A somewhat different way to say this is, the more components of the equation you can include in the approximation the better. In any case, our conclusion is that  $\alpha = \gamma = 2$  and this yields the next problem to solve.

$$O(\varepsilon^3) \quad \frac{\partial^2}{\partial t_1^2} \theta_1 + \theta_1 + 2 \frac{\partial^2}{\partial t_1 \partial t_2} \theta_0 = \frac{1}{6} \theta_0^3$$

$$\theta_1(0, 0) = 0, \quad \frac{\partial}{\partial t_1} \theta_1(0, 0) + \frac{\partial}{\partial t_2} \theta_0(0, 0) = 0$$

The method of undetermined coefficients can be used to find a particular solution. To be able to do this we first substitute the solution for  $\theta_0$  into the differential equation and then use the identity  $\cos^3(t) = \frac{1}{4}(3 \cos(t) + \cos(3t))$ . The result is

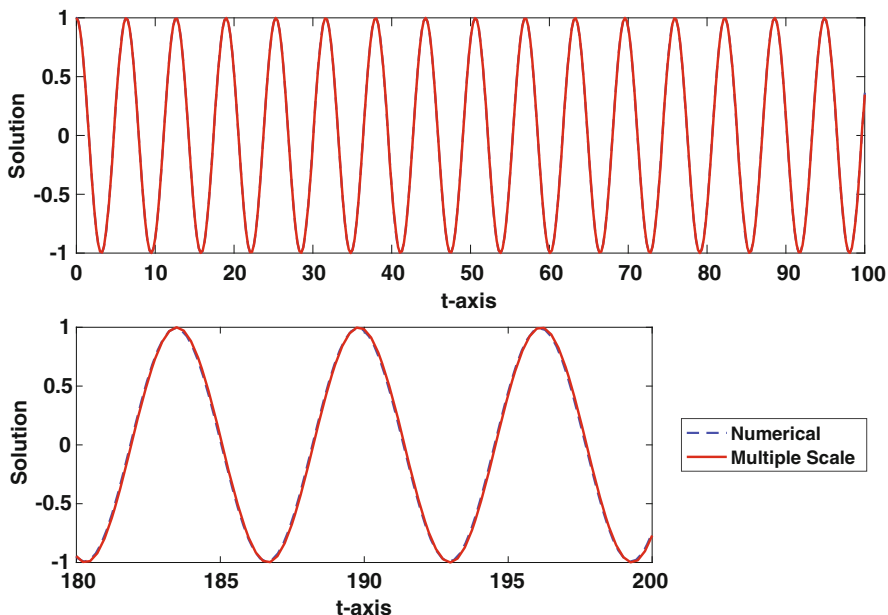
$$\begin{aligned} \theta_1'' + \theta_1 = & \frac{1}{24} [3 \cos(t_1 + B) + \cos(3(t_1 + B))] \\ & + 2A' \sin(t_1 + B) + 2AB' \cos(t_1 + B). \end{aligned} \quad (2.110)$$

We are at a similar point to what occurred using a regular expansion, as given in (2.100). As before, the right-hand side of the differential equation contains functions that are solutions of the associated homogeneous equation, namely,  $\cos(t_1 + B)$  and  $\sin(t_1 + B)$ . If they are allowed to remain they will produce a solution containing either  $t_1 \cos(t_1 + B)$  or  $t_1 \sin(t_1 + B)$ . Either one will cause a secular term in the expansion and for this reason we will select  $A$  and  $B$  to prevent this from happening. To lose  $\sin(t_1 + B)$  we take  $A' = 0$  and to eliminate  $\cos(t_1 + B)$  we take  $2AB' = -\frac{1}{8}$ . With the earlier determined initial conditions  $A(0) = 1$  and  $B(0) = 0$ , we conclude that  $A = 1$  and  $B = -t_2/16$ .

In the above analysis we never actually determined  $\theta_1$ . It is enough to know that the problem for  $\theta_1$  will not result in a secular term in the expansion. We did find  $A$  and  $B$ , and with them the expansion is

$$\theta \sim \varepsilon \cos(t - \varepsilon^2 t/16) + \dots \quad (2.111)$$

To investigate the accuracy of this approximation, it is plotted in Fig. 2.23 using the same values as for Fig. 2.22. The numerical solution is also plotted in both graphs, and it is hard to distinguish it from the multiple scale approximation curve. Clearly we have improved the first term approximation, and now do well with both amplitude and phase.



**Fig. 2.23** Graph of the numerical solution of the pendulum problem (2.80)–(2.82) and the multiple scale approximation (2.111). Shown are the solutions for  $0 \leq t \leq 100$ , as well as a close up of the solutions near  $t = 200$ . In the calculation  $\varepsilon = \frac{1}{3}$  and the solution has been divided by  $\varepsilon = \frac{1}{3}$

## Exercises

### Sections 2.1 and 2.2

**2.1** Find the first three terms in the asymptotic expansion of each function. Your answer should be of the form  $f \sim a_1 \varepsilon^\alpha + a_2 \varepsilon^\beta + a_3 \varepsilon^\gamma$ , where  $\alpha < \beta < \gamma$ , and  $a_1, a_2, a_3$  are nonzero.

- |                                                                                   |                                                      |
|-----------------------------------------------------------------------------------|------------------------------------------------------|
| (a) $f = (1 + 6\varepsilon)/(1 + \varepsilon^2)$ .                                | (i) $f = 1/(1 - e^\varepsilon)$ .                    |
| (b) $f = [1 + \cos(\varepsilon)]^5$ .                                             | (j) $f = 1/\ln(1 + \varepsilon)$ .                   |
| (c) $f = \sin(\frac{\pi}{2} + \varepsilon)$ .                                     | (k) $f = \sqrt{\varepsilon} + \cos(\varepsilon)$ .   |
| (d) $f = \sqrt{(1 - \varepsilon)/(1 + \varepsilon)}$ .                            | (l) $f = \sin(\sqrt{\varepsilon})$ .                 |
| (e) $f = \sin(\varepsilon)/(1 + \varepsilon)$ .                                   | (m) $f = \varepsilon^{1/3} \ln(1 + \varepsilon^2)$ . |
| (f) $f = (e^\varepsilon - e^{-\varepsilon})/(e^\varepsilon + e^{-\varepsilon})$ . | (n) $f = \sqrt{\varepsilon} \cos(\varepsilon)$ .     |
| (g) $f = \sqrt{\varepsilon + 3\varepsilon^3}$ .                                   | (o) $f = e^{\sqrt{\varepsilon}}$ .                   |
| (h) $f = 1/\sqrt{\sin(\varepsilon)}$ .                                            | (p) $f = \sqrt{\varepsilon}/(1 + \varepsilon)$ .     |

**2.2** Consider the equation

$$x^3 + \varepsilon x - 8 = 0.$$

- Sketch the functions in this equation and then use this to explain why there is one real-valued solution.
- Find a two-term asymptotic expansion, for small  $\varepsilon$ , of the solution.
- Find a three-term asymptotic expansion, for small  $\varepsilon$ , of each solution.

**2.3** Consider the equation

$$1 - x^2 = \varepsilon e^{-x}.$$

- Sketch the functions in this equation and then use this to explain why there are two solutions and describe where they are located for small values of  $\varepsilon$ .
- Find a two-term asymptotic expansion, for small  $\varepsilon$ , of each solution.
- Find a three-term asymptotic expansion, for small  $\varepsilon$ , of each solution.

**2.4** Consider the equation

$$\sin(\varepsilon x) = 1 - 2x.$$

- Sketch the functions in this equation and then use this to explain why there is one solution.

- (b) Find a two-term asymptotic expansion, for small  $\varepsilon$ , of the solution.
- (c) Find a three-term asymptotic expansion, for small  $\varepsilon$ , of each solution.

**2.5** Consider the equation

$$\sqrt{1+x} = \varepsilon(4-x^2).$$

- (a) Sketch the functions in this equation and then use this to explain why there is one solution.
- (b) Find a two-term asymptotic expansion, for small  $\varepsilon$ , of the solution.

**2.6** Consider the equation

$$\ln x + \varepsilon x = 4.$$

- (a) Sketch the functions in this equation and then use this to explain why there is only one solution and describe where it is located for small values of  $\varepsilon$ .
- (b) Find a two-term asymptotic expansion, for small  $\varepsilon$ , of the solution.

**2.7** Consider the equation

$$xe^x = \varepsilon.$$

- (a) Sketch the functions in this equation and then use this to explain why there is one solution.
- (b) Find a two-term asymptotic expansion, for small  $\varepsilon$ , of the solution.

**2.8** Consider the equation

$$x^2 = \varepsilon \cos x.$$

- (a) Sketch the functions in this equation and then use this to explain why there are two solutions and describe where they are located for small values of  $\varepsilon$ .
- (b) Find a two-term asymptotic expansion, for small  $\varepsilon$ , of each solution.

**2.9** Consider the equation

$$\frac{x^3}{1+x} = \varepsilon.$$

- (a) Sketch the functions in this equation and then use this to explain why there is only one solution and describe where it is located for small values of  $\varepsilon$ .
- (b) Find a two-term asymptotic expansion, for small  $\varepsilon$ , of the solution.

**2.10** The nondimensional form of the Newton-Sefan law of cooling is (see Exercise 1.20)

$$\frac{du}{d\tau} = -u - \frac{\alpha}{\varepsilon}[(1 + \varepsilon u)^4 - 1],$$

where  $u(0) = 1$  and  $\alpha$  is a positive constant.

- (a) Find a two-term expansion of  $u$  for small  $\varepsilon$ .
- (b) Transform back into dimensional variables and obtain an expansion for  $T(t)$ .

**2.11** For the Michaelis-Menten system, which is considered in Sect. 3.3.1, one comes across the problem of solving

$$\frac{du}{dt} = -\frac{\gamma u}{1 + \varepsilon u},$$

where  $u(0) = 1$  and  $\gamma$  is a positive constant.

- (a) For small  $\varepsilon$ , find a two-term expansion of the solution.
- (b) Solve the IVP and show that the solution satisfies  $\ln u + \varepsilon u = -\gamma t + \varepsilon$ . This is an example of an implicitly defined solution, which are very common when solving nonlinear differential equations.
- (c) Using the result from part (b), find a two-term expansion of  $u$  for small  $\varepsilon$ .

**2.12** As shown in Exercise 1.19, the nondimensional problem for the velocity of a sphere dropping through the atmosphere satisfies

$$\frac{du}{d\tau} = -1 - u + \varepsilon u^2,$$

where  $u(0) = 0$ .

- (a) For small  $\varepsilon$ , find a two-term expansion of the solution.
- (b) Solve the IVP for  $u(\tau)$ , and then derive a two-term expansion of it for small  $\varepsilon$ .
- (c) Transform back into dimensional variables and obtain an expansion for  $v(t)$ .

**2.13** The projectile problem that includes air resistance is

$$\frac{d^2x}{dt^2} + \frac{dx}{dt} = -\frac{1}{(1 + \varepsilon x)^2},$$

where  $x(0) = 0$ , and  $\frac{dx}{dt}(0) = 1$ . For small  $\varepsilon$ , find a two-term expansion of the solution.

**2.14** Air resistance is known to depend nonlinearly on velocity, and the dependence is often assumed to be quadratic. Assuming gravity is constant, the equations of motion are



$$\frac{d^2 y}{dt^2} = -\varepsilon \frac{dy}{dt} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2},$$

$$\frac{d^2 x}{dt^2} = -1 - \varepsilon \frac{dx}{dt} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}.$$

Here  $x$  is the vertical elevation of the object, and  $y$  is its horizontal location. The initial conditions are  $x(0) = y(0) = 0$ , and  $\frac{dx}{dt}(0) = \frac{dy}{dt}(0) = 1$ . The assumption is that air resistance is weak, and so  $\varepsilon$  is small and positive.

- (a) For small  $\varepsilon$ , find the first terms in the expansions for  $x$  and  $y$ .  
 (b) Find the second terms in the expansions for  $x$  and  $y$ .

**2.15** Consider the nonlinear boundary value problem

$$\frac{d}{dx} \left( \frac{y_x}{1 + \varepsilon y_x^2} \right) - y = 0, \quad \text{for } 0 < x < 1,$$

where  $y(0) = 1$  and  $y(1) = e^{-1}$ . This type of nonlinearity arises in elasticity, a topic taken up in Chap. 6. For small  $\varepsilon$ , find a two-term expansion of the solution.

## Section 2.3

**2.16** The following are scale functions. Determine which is  $\phi_1$ ,  $\phi_2$ , etc.

- (a)  $2\varepsilon, \varepsilon^3, 5$ .  
 (b)  $1 + \varepsilon, \varepsilon^2, \varepsilon/(1 + \varepsilon)$ .  
 (c)  $\varepsilon, e^\varepsilon, \varepsilon \sin \varepsilon$ .  
 (d)  $\varepsilon^{-1}, \varepsilon^{1/2}, \varepsilon^{-1/3}, \varepsilon^{1/4}$ .  
 (e)  $e^\varepsilon - 1 - \varepsilon, e^\varepsilon - 1, e^\varepsilon, e^\varepsilon - 1 - \varepsilon - \frac{1}{2}\varepsilon^2$ .

**2.17** How small  $\varepsilon$  must be to obtain an accurate numerical approximation depends on what scale functions you have. In the following show that  $\phi_2 \ll \phi_1$ . Also, assuming that  $y \sim \phi_1 + \phi_2$ , how small does  $\varepsilon$  have to be so that  $\phi_2$  is no more than one-tenth of  $\phi_1$ ?

- (a)  $\phi_1 = \varepsilon, \phi_2 = \varepsilon^2$ .  
 (b)  $\phi_1 = e^{-1/\varepsilon}, \phi_2 = e^{-2/\varepsilon}$ .  
 (c)  $\phi_1 = -\ln \varepsilon, \phi_2 = \ln(-\ln \varepsilon)$ .

It is worth noting that these scale functions arise in the expansion for the large  $\beta$  solution in Exercise 2.39(b).

**2.18** This problem concerns showing that (2.12) qualifies as an asymptotic expansion.

- What are the three scales functions  $\phi_1$ ,  $\phi_2$ ,  $\phi_3$  appearing in (2.12)? What are their respective coefficients  $f_1$ ,  $f_2$ ,  $f_3$ ?
- Show that  $f \sim f_1\phi_1$  is a one term asymptotic expansion by showing that (2.37) holds.
- Show that  $f \sim f_1\phi_1 + f_2\phi_2$  is a two term asymptotic expansion by showing that (2.38) holds.
- Show that  $f \sim f_1\phi_1 + f_2\phi_2 + f_3\phi_3$  is a three term asymptotic expansion by writing down the limit requirement, and then showing that it is satisfied by (2.12).

## Sections 2.4–2.6

**2.19** Consider the equation  $\varepsilon x^3 - 2x + 1 = 0$ .

- Sketch the functions in this equation and then use this to describe where the three solutions of the equation are located for small values of  $\varepsilon$ .
- Find a two-term asymptotic expansion, for small  $\varepsilon$ , of each solution.

**2.20** Consider the equation  $\varepsilon x^4 - x - 1 = 0$ .

- Sketch the functions in this equation and then use this to explain why there are only two solutions to this equation and describe where they are located for small values of  $\varepsilon$ .
- Find a two-term asymptotic expansion, for small  $\varepsilon$ , of each solution.

**2.21** Consider the equation

$$\frac{1-x}{1+x} = \varepsilon x^3.$$

- Sketch the functions in this equation and then use this to explain why there are two solutions and describe where they are located for small values of  $\varepsilon$ .
- Find a two-term asymptotic expansion, for small  $\varepsilon$ , of the solutions.

**2.22** The Friedrichs' (1942) model problem for a boundary layer in a viscous fluid is

$$\varepsilon y'' = a - y', \quad \text{for } 0 < x < 1,$$

where  $y(0) = 0$  and  $y(1) = 1$  and  $a$  is a given positive constant. The boundary layer in this problem is at  $x = 0$ .

- After finding a first term of the inner and outer expansions, derive a composite expansion of the solution.

- (b) Taking  $a = 1$ , plot the exact and composite solutions, on the same axes, for  $\varepsilon = 10^{-1}$ . Do the same thing for  $\varepsilon = 10^{-2}$  and for  $\varepsilon = 10^{-3}$ . Comment on the effectiveness, or noneffectiveness, of the expansion in approximating the solution.
- (c) Suppose you assume the boundary layer is at the other end of the interval. Show that the resulting first term approximations from the inner and outer regions do not match.

**2.23** The following problem has a boundary layer at  $x = 0$ :

$$\varepsilon y'' + 3y' - y^4 = 0, \quad \text{for } 0 < x < 1,$$

where  $y(0) = 1$  and  $y(1) = 1$ .

- (a) Find a composite expansion of the solution.  
 (b) Sketch the solution.

**2.24** The following problem has a boundary layer at  $x = 0$ :

$$\varepsilon^2 y'' + y' + \varepsilon y = x, \quad \text{for } 0 < x < 1,$$

where  $y(0) = 1$  and  $y(1) = 1$ .

- (a) Find a composite expansion of the solution.  
 (b) Sketch the solution.

**2.25** The following problem has a boundary layer at  $x = 0$ :

$$\varepsilon y'' + y' - y = 0, \quad \text{for } 0 < x < 1,$$

where  $y(0) = 0$ ,  $y(1) = -1$ .

- (a) Find a composite expansion of the solution.  
 (b) Sketch the solution.

**2.26** The following problem has a boundary layer at  $x = 0$ :

$$\varepsilon y'' + 2y' - y^3 = 0, \quad \text{for } 0 < x < 1,$$

where  $y(0) = 0$  and  $y(1) = 1$ .

- (a) Find a composite expansion of the solution.  
 (b) Sketch the solution.

**2.27** The following problem has a boundary layer at  $x = 0$ :

$$\varepsilon y'' + y' + \varepsilon y = 0, \quad \text{for } 0 < x < 1,$$

where  $y(0) = 1$  and  $y(1) = 2$ .

- (a) Find a composite expansion of the solution.  
 (b) Sketch the solution.

**2.28** The following problem has a boundary layer at  $x = 1$ :

$$\varepsilon y'' - 3y' - y^4 = 0, \quad \text{for } 0 < x < 1,$$

where  $y(0) = 1$  and  $y(1) = 1$ .

- (a) Find a composite expansion of the solution.
- (b) Sketch the solution.

**2.29** The following problem has a boundary layer at  $x = 1$ :

$$\varepsilon y'' - \frac{1}{2}y' - xy = 0, \quad \text{for } 0 < x < 1,$$

where  $y(0) = 1$  and  $y(1) = 1$ .

- (a) Find a composite expansion of the solution.
- (b) Sketch the solution.

**2.30** The following problem has a boundary layer at each end of the interval:

$$\varepsilon y'' - (1 + 3x)y = -1, \quad \text{for } 0 < x < 1,$$

where  $y(0) = 0$  and  $y(1) = 2$ .

- (a) Find a composite expansion of the solution.
- (b) Sketch the solution.

**2.31** The following problem has a boundary layer at each end of the interval:

$$\varepsilon y'' + \varepsilon y' - (1 + x)y = 2, \quad \text{for } 0 < x < 1,$$

where  $y(0) = 0$  and  $y(1) = 0$ .

- (a) Find a composite expansion of the solution.
- (b) Sketch the solution.

## Section 2.7

**2.32** As found in Exercise 1.18, the equation for a weakly damped oscillator is

$$y'' + \varepsilon y' + y = 0, \quad \text{for } 0 < t,$$

where  $y(0) = 1$  and  $y'(0) = 0$ .

- (a) For small  $\varepsilon$ , find a two-term regular expansion of the solution.
- (b) Explain why the expansion in (a) is not well ordered for  $0 \leq t < \infty$ . What requirement is needed on  $t$  so it is well ordered?
- (c) Use two-timing to construct a better approximation to the solution.

**2.33** The weakly nonlinear Duffing equation is

$$y'' + y + \varepsilon y^3 = 0, \quad \text{for } 0 < t,$$

where  $y(0) = 0$  and  $y'(0) = 1$ .

- For small  $\varepsilon$ , find a two-term regular expansion of the solution.
- Explain why the expansion in (a) is not well ordered for  $0 \leq t < \infty$ . What requirement is needed on  $t$  so it is well ordered?
- Use two-timing to construct a better approximation to the solution.

### *Additional Questions*

**2.34** To determine the displacement  $u(x)$  of a nonlinear bunge cord of length  $\ell$  it is necessary to solve

$$E \frac{du}{dx} + K \left( \frac{du}{dx} \right)^3 = \rho g(\ell - x), \quad \text{for } 0 < x < \ell,$$

where  $u(0) = 0$ . Here  $E$  is the elastic modulus,  $\rho$  is the mass density, and  $g$  is the gravitational acceleration constant. This equation is derived in Sect. 6.8.3.

- What are the dimensions of  $K$ ?
- Nondimensionalize the problem taking  $u(x) = u_c v(s)$  and  $x = x_c s$ . The resulting problem should have the form

$$\frac{dv}{ds} + \varepsilon \left( \frac{dv}{ds} \right)^3 = 1 - s, \quad \text{for } 0 < s < 1,$$

where  $v(0) = 0$ .

- Find the first two terms in the expansion of  $v$ , and then transform this back into dimensional variables and obtain an expansion for  $u(x)$ .

**2.35** An equation for the displacement  $u(x)$  of an elastic string due to gravity is (Chen and Ding, 2008)

$$T \left[ 1 + \left( \frac{du}{dx} \right)^2 \right] \frac{d^2 u}{dx^2} = -\rho g, \quad \text{for } 0 < x < \ell,$$

where  $u(0) = 0$  and  $u(\ell) = 0$ . Also,  $T$  is the tension,  $\rho$  is the mass density per unit length, and  $g$  is the gravitational acceleration constant.

- Nondimensionalize the problem taking  $u(x) = u_c v(s)$  and  $x = x_c s$ . The resulting problem should have the form

$$\left[ 1 + \varepsilon \left( \frac{dv}{ds} \right)^2 \right] \frac{d^2 v}{ds^2} = -1, \quad \text{for } 0 < s < 1,$$

where  $v(0) = 0$  and  $v(1) = 0$ .

- (b) It is going to be assumed that  $\varepsilon$  is small. What is this assuming about the tension  $T$ ?
- (c) Find the first two terms in the expansion of  $v$ .
- (d) Writing the expansion as  $v \sim v_0(s) + \varepsilon v_1(s)$ , on the same axes sketch  $v_0(s)$  and  $v_0(s) + \varepsilon v_1(s)$ . Does the nonlinearity in the differential equation cause the string to be more, or less, resistant to stretching?

**2.36** The concentration of a chemical species is known to satisfy what is called a reaction-diffusion equation (these are derived in Chap. 4). From this, it is found that the steady-state concentration  $u(x)$  of the species satisfies

$$D \frac{d^2 u}{dx^2} + c \frac{du}{dx} - ku + \gamma = 0, \text{ for } 0 < x < \ell,$$

where  $u(0) = 0$  and  $u(\ell) = 0$ . Also,  $D$  is the diffusion coefficient. Assume that all of the coefficients ( $D$ ,  $c$ ,  $k$ , and  $\gamma$ ) are positive constants.

- (a) Nondimensionalize the problem taking  $u(x) = u_c v(s)$  and  $x = x_c s$ . The resulting problem should have the form

$$\varepsilon \frac{d^2 v}{ds^2} + \alpha \frac{dv}{ds} - v + 1 = 0, \text{ for } 0 < s < 1,$$

where  $v(0) = 0$  and  $v(1) = 0$ .

- (b) It is going to be assumed that  $\varepsilon$  is small. What is this assuming about the diffusion coefficient  $D$ ?
- (c) Given that there is a boundary layer at the left end, find a composite expansion of  $v$ .
- (d) Transform your composite expansion from part (c) back into dimensional variables and obtain an expansion for  $u(x)$ .

**2.37** This problem derives additional information from the projectile problem.

- (a) Let  $s_M$  be the (nondimensional) time at which the projectile reaches its maximum height. Given that the solution depends on  $\varepsilon$ , it follows that  $s_M$  depends on  $\varepsilon$ . Use (2.29) to find a two-term expansion of  $s_M$  for small  $\varepsilon$ . What is the resulting two-term expansion for the maximum height  $u_M$ ?
- (b) Let  $s_E$  be the (nondimensional) time at which the projectile hits the ground. Given that the solution depends on  $\varepsilon$ , it follows that  $s_E$  depends on  $\varepsilon$ . Use (2.29) to find a two-term expansion of  $s_E$  for small  $\varepsilon$ .
- (c) Based on your results from parts (a) and (b), describe the effects of the nonlinear gravitational field on the motion of the projectile.

**2.38** In this problem assume that

$$x \sim x_0 \varepsilon^\alpha + x_1 \varepsilon^\beta + x_2 \varepsilon^\gamma + \cdots,$$

where  $\alpha < \beta < \gamma < \dots$ . Also assume that the  $x_i$ 's are nonzero.

(a) Show that the resulting two-term expansion of  $x^n$  is

$$x^n \sim x_0^n \varepsilon^{n\alpha} + nx_1 x_0^{n-1} \varepsilon^{\beta+(n-1)\alpha}.$$

(b) If  $\gamma = 2\beta - \alpha$ , then show that the resulting three-term expansion of  $x^n$  is

$$x^n \sim x_0^n \varepsilon^{n\alpha} + nx_1 x_0^{n-1} \varepsilon^{\beta+(n-1)\alpha} + nx_0^{n-2} \left[ x_0 x_2 + \frac{1}{2}(n-1)x_1^2 \right] \varepsilon^{2\beta+(n-2)\alpha}.$$

(c) What is the resulting three-term expansion of  $x^n$  when  $\gamma > 2\beta - \alpha$ ?

(d) What is the resulting three-term expansion of  $x^n$  when  $\beta < \gamma < 2\beta - \alpha$ ?

**2.39** In the study of reactions of chemical mixtures one comes across the following problem:

$$\frac{d^2 y}{dx^2} = -\varepsilon e^y, \quad \text{for } 0 < x < 1,$$

where  $y(0) = y(1) = 0$ . This is known as Bratu's equation, and it illustrates some of the difficulties one faces when solving nonlinear equations.

(a) Explain why a boundary layer is not expected in the problem and find the first two terms in a regular expansion of the solution.

(b) The function

$$y = -2 \ln \left[ \frac{\cosh(\beta(1-2x))}{\cosh(\beta)} \right],$$

where  $\beta$  satisfies

$$\cosh(\beta) = 2\beta \sqrt{\frac{2}{\varepsilon}}, \quad (2.112)$$

is a solution of the Bratu problem. By sketching the functions in (2.112), as functions of  $\beta$ , explain why there is an  $\varepsilon_0$  where if  $0 < \varepsilon < \varepsilon_0$ , then there are exactly two solutions, while if  $\varepsilon_0 < \varepsilon$ , then there are no solutions.

(c) Which of the two solutions in part (b), if any, corresponds to the one you found in part (a)? Note that to answer this you need to expand the solution in part (b).

(d) Comment on the conclusion drawn from part (b) and your result in part (a). Explain why the regular expansion does not fail in a manner found in a boundary layer problem but that it is still not adequate for this problem.

## Chapter 3

### Kinetics

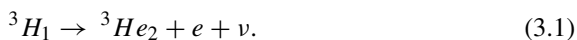


### 3.1 Introduction

We now investigate how to model, and analyze, the interactions of multiple species and how these interactions produce changes in their populations. Examples of such problems are below.

#### 3.1.1 Radioactive Decay

A radioactive isotope is unstable and will decay by emitting a particle, transforming into another isotope. As an example, tritium  ${}^3H_1$  is a radioactive form of hydrogen that occurs when cosmic rays interact with the atmosphere. It decays by emitting an electron  $e$  and antineutrino  $\nu$  to produce a stable helium isotope  ${}^3He_2$ . The conventional way to express this conversion is



The assumption used to model such situations is that the rate of decrease in the amount of radioactive isotope is proportional to the amount currently present. To translate this into mathematical terms, let  $N(t)$  designate the amount of the radioactive material present at time  $t$ . In this case we obtain the rate equation

$$\frac{dN}{dt} = -kN, \quad \text{for } 0 < t, \quad (3.2)$$



where

$$N(0) = N_0. \quad (3.3)$$

In the above equation  $k$  is the proportionality constant and it is assumed to be positive.

### 3.1.2 Predator-Prey

This involves two species and a typical situation is a population of predators, wolves, which survives by eating another species, rabbits. To write down a model for their interaction, let  $R(t)$  and  $W(t)$  denote the number of rabbits and wolves, respectively. In this case, we have

$$\frac{dR}{dt} = aR - bRW, \quad (3.4)$$

$$\frac{dW}{dt} = -cW + dRW. \quad (3.5)$$

In the above equations  $a$ ,  $b$ ,  $c$ , and  $d$  are proportionality constants. To obtain the first equation, it has been assumed that the population of rabbits, with wolves absent, increases at a rate proportional to their current population ( $aR$ ). When the wolves are present it is assumed the rabbit population decreases at a rate proportional to both populations ( $-bRW$ ). Similarly, for the second equation, the number of wolves, with rabbits absent, decreases at a rate proportional to their current population ( $-cW$ ), but increases at a rate proportional to both the rabbit and wolf populations when rabbits are available ( $dRW$ ). To complete the formulation we need the initial concentrations, given as  $R(0) = R_0$  and  $W(0) = W_0$ .

### 3.1.3 Epidemic Model

Epidemics, such as the black death and cholera, have come and gone throughout human history. Given the catastrophic nature of these events there is a long history of scientific study trying to predict how and why they occur. One of particular prominence is the Kermack-McKendrick model for epidemics. This assumes the population can be separated into three groups. One is the population  $S(t)$  of those susceptible to the disease, another is the population  $I(t)$  that is ill, and the third is the population  $R(t)$  of individuals that have recovered. A model that accounts for the susceptible group getting sick, the subsequent increase in the ill population, and the eventual increase in the recovered population is the following set of equations:

$$\begin{aligned}\frac{dS}{dt} &= -k_1 SI, \\ \frac{dI}{dt} &= -k_2 I + k_1 SI, \\ \frac{dR}{dt} &= k_2 I,\end{aligned}$$

where  $S(0) = S_0$ ,  $I(0) = I_0$ , and  $R(0) = R_0$ . In the above equations  $k_1, k_2$  are proportionality constants. Given the three groups, and the letters used to designate them, this is an example of what is known as a SIR model in mathematical epidemiology. This model does not account for births or deaths, and for this reason the total population stays constant. This can be seen in the above equations because

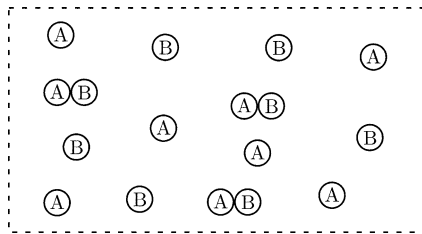
$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0,$$

or in other words  $\frac{d}{dt}(S + I + R) = 0$ . The fact that  $S + I + R$  is constant is an example of a conservation law, and these will play a prominent role in this chapter.

## 3.2 Kinetic Equations

The common thread in the above examples is that one or more species combine, or transform, to form new or additional species. This is a situation common in chemistry and we will extend the theory developed in chemical kinetics to describe interacting populations or species. The main result is the Law of Mass Action and to motivate how it is derived consider a region containing a large number of two species, labeled as  $A$  and  $B$ . A small portion of this region is shown in Fig. 3.1. As indicated in the figure, both species are assumed to be distributed throughout the region. It is also assumed that they are in motion, and when an  $A$  and  $B$  come into contact they can combine to form a new species  $C$ . The  $C$ 's are shown in the figure with an  $A$  and  $B$  stuck together. The symbolism for this is

**Fig. 3.1** Sample domain illustrating assumptions underlying the Law of Mass Action, where two species combine to form a third





The question is, can we use this information to determine the concentrations of the three species?

The reaction in (3.6) states that one  $A$  and one  $B$  are used to construct one  $C$ . This means that the rate of change of the concentrations of  $A$  and  $B$  is the same, and they are the negative of the change in  $C$ . In other words,

$$\frac{dA}{dt} = \frac{dB}{dt} = -\frac{dC}{dt}. \quad (3.7)$$

In the above expressions there is a mild case of notation abuse in the sense that we are letting  $A$ ,  $B$ , and  $C$  also designate the concentrations of the respective species. This dual usage of having the letters designate individual molecules as in (3.6), and concentrations as in (3.7), is common in kinetics and should not cause problems in the development.

The equalities in (3.7) can be rewritten as

$$\frac{dA}{dt} = -r, \quad (3.8)$$

$$\frac{dB}{dt} = -r, \quad (3.9)$$

$$\frac{dC}{dt} = r, \quad (3.10)$$

where  $r$  is known as the *rate of the reaction*. The Law of Mass Action will be used to determine  $r$ . This will involve several assumptions, and the discussion below will help in understanding the reasons for some of them.

It is reasonable to assume that  $r$  depends on the collision frequency of  $A$  and  $B$ , and this means it depends on the concentrations of  $A$  and  $B$ . Also, if there are no  $A$ 's, or if there are no  $B$ 's, then the rate is zero. We are therefore assuming that  $r = r(A, B)$ , where  $r(A, 0) = r(0, B) = 0$ . To obtain a first term approximation of this function we use Taylor's theorem to obtain

$$r = r_{00} + r_{10}A + r_{01}B + r_{20}A^2 + r_{11}AB + r_{02}B^2 + \dots$$

In this expression (see Sect. A.2)

$$r_{00} = r(0, 0),$$

$$r_{10} = \frac{\partial r}{\partial A}(0, 0), \quad r_{01} = \frac{\partial r}{\partial B}(0, 0),$$

$$r_{20} = \frac{1}{2} \frac{\partial^2 r}{\partial A^2}(0, 0), \quad r_{02} = \frac{1}{2} \frac{\partial^2 r}{\partial B^2}(0, 0).$$

All of these terms are zero. For example, because  $r(A, 0) = 0$ , for any value of  $A$ , it follows that

$$\frac{\partial r}{\partial A}(A, 0) = 0 \quad \text{and} \quad \frac{\partial^2 r}{\partial A^2}(A, 0) = 0.$$

Similarly, because  $r(0, B) = 0$ , it follows that  $r_{01} = r_{02} = 0$ . What is not necessarily zero is the mixed derivative term

$$r_{11} = \frac{\partial^2 r}{\partial A \partial B}(0, 0).$$

Therefore, the first nonzero term in the Taylor series is  $r_{11}AB$ , and from this we have

$$r = kAB, \tag{3.11}$$

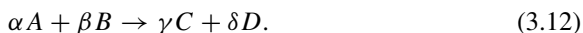
where  $k$  is known as the rate constant, and it is required to be positive. This expression, along with the rate equations in (3.8)–(3.10), is the Law of Mass Action as applied to the reaction in (3.6). This will be generalized to more complicated reactions in the next section.

A variation of the above reaction is the case of when  $A$  and  $B$  are the same species. In this case (3.6) becomes  $2A \rightarrow C$  and (3.11) takes the form  $r = kA^2$ . Also, we no longer have an equation for  $B$ , but because two  $A$ 's are now lost every time a  $C$  is produced, then (3.8) becomes  $A' = -2r$ . The equation for  $C$  stays the same. This shows that the coefficients in the reaction play a role as multiplicative factors in the rate equation, as well as in the formula for  $r$ .

Finally, for the reactions to take place the reactants must be moving to find each other. It's implicitly assumed in what follows that this is occurring, and that the concentrations of the reactants do not depend on their spatial locations. This is often referred to as a *well-stirred*, or *well-mixed*, assumption. In the next chapter we will consider how to incorporate spatial dependence into the problem, yielding what are called reaction-diffusion equations.

### 3.2.1 The Law of Mass Action

To state the general form of the Law of Mass Action certain terms need to be defined. For this we generalize the above example and consider the reaction



The coefficients  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$  are positive constants known as the *stoichiometric coefficients* for the reaction. In effect, this reaction states that  $\alpha$  of the  $A$ 's combine

with  $\beta$  of the  $B$ 's to form  $\gamma$  of the  $C$ 's and  $\delta$  of the  $D$ 's. Said this way, the implication is that the stoichiometric coefficients are integers. The fact is that they generally are, although we will not make this assumption explicitly. The species on the left,  $A$  and  $B$ , are the *reactants* and those on the right,  $C$  and  $D$ , are the *products* for this particular reaction. The order of the reaction is the total number of reactants, which in this case is  $\alpha + \beta$ .

The Law of Mass Action, which will be stated shortly, states that the rate  $r$  of the reaction in (3.12) is

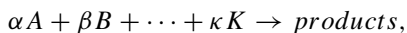
$$r = kA^\alpha B^\beta, \quad (3.13)$$

where  $k$  is the rate constant, or the reaction rate coefficient, and it is required to be positive. In writing down this formula the notation has been corrupted a bit. As happened in the earlier example, we started off letting  $A$  and  $B$  designate the reactants but in the rate formula (3.13) these same letters have been used to designate their concentrations.

We are now in position to state the assumptions underlying the Law of Mass Action.

**Law of Mass Action.** *This consists of the following three assumptions:*

1. *For the reaction*



*the rate  $r$  is*

$$r = kA^\alpha B^\beta \cdots K^\kappa,$$

*where  $k$  is the rate constant.*

2. *The rate of change of the concentration of a species  $X$ , with stoichiometric coefficient  $\nu$ , is*

$$\frac{d}{dt}X = \pm \nu r,$$

*where the  $+$  is used when  $X$  is a product and the  $-$  when  $X$  is a reactant.*

3. *For a system of reactions, the rates add.*

As an example, for the reaction in (3.12), the rate of change  $\frac{dA}{dt}$  is equal to  $-\alpha r$ , while the rate of change  $\frac{dC}{dt}$  is equal to  $\gamma r$ . Combining this information, from the Law of Mass Action the kinetic equations for the concentrations are

$$\begin{aligned} \frac{dA}{dt} &= -\alpha r \\ &= -\alpha k A^\alpha B^\beta, \end{aligned} \quad (3.14)$$

$$\begin{aligned}\frac{dB}{dt} &= -\beta r = -\beta k A^\alpha B^\beta, \\ \frac{dC}{dt} &= \gamma r = \gamma k A^\alpha B^\beta, \\ \frac{dD}{dt} &= \delta r = \delta k A^\alpha B^\beta.\end{aligned}$$

To complete the formulation, it is assumed that the initial concentrations are known, and so,  $A(0) = A_0$ ,  $B(0) = B_0$ ,  $C(0) = C_0$ ,  $D(0) = D_0$  are given.

The specific units of the terms in the above equations depend on the application. For example, if the species are chemicals, then concentration, using SI units, is measured in moles per decimeter ( $\text{mol/dm}^3$ ). It is not unusual, however, to find that when using liquids that concentrations are measured using molarity ( $M$ ) where  $1M = 6.02 \times 10^{23}$  molecules per liter. In applications involving gases the units that are often used are moles per cubic centimeter. If the application involves populations, then population density (e.g., number per area) is usually used. Whatever the application, the units for the rate constant depend on the specific reaction. This can be seen from (3.14) because  $[A'] = [k][A^\alpha B^\beta]$ . If  $A$  and  $B$  are concentrations, then  $[k] = T^{-1} L^{3(\alpha+\beta-1)}$ . Consequently, the units for the rate coefficient for  $A + B \rightarrow C$  are different than they are for the reaction  $A + 2B \rightarrow C$ .

### 3.2.2 Conservation Laws

We have produced four equations for the four species involved in the example reaction in (3.12). Although they are not particularly easy to solve there is one significant simplification we are able to make. To explain what this is, note that it is possible to combine the first two equations to produce zero on the right-hand side. Specifically,  $\frac{d}{dt}(\beta A - \alpha B) = 0$  and this means  $\beta A - \alpha B = \text{constant}$ . Using the stated initial conditions it follows that

$$\beta A - \alpha B = \beta A_0 - \alpha B_0. \quad (3.15)$$

In a similar manner, by combining the  $C$  and  $A$  equations we obtain

$$\gamma A - \alpha C = \gamma A_0 - \alpha C_0, \quad (3.16)$$

and from the  $D$  and  $A$  equations

$$\delta A - \alpha D = \delta A_0 - \alpha D_0. \quad (3.17)$$

Equations (3.15)–(3.17) are conservation laws. These will play an essential role in our study of kinetic equations, so it is important to define exactly what this means.

**Conservation Law.** Given species  $A, B, C, \dots, Z$  and nonzero numbers  $a, b, c, \dots, z$ , then  $aA + bB + cC + \dots + zZ$  is said to be conserved if

$$\frac{d}{dt}(aA + bB + cC + \dots + zZ) = 0. \quad (3.18)$$

The corresponding conservation law is

$$aA + bB + cC + \dots + zZ = aA_0 + bB_0 + cC_0 + \dots + zZ_0, \quad (3.19)$$

where  $A_0, B_0, C_0, \dots, Z_0$  are the initial values for the respective species.

A particularly useful application of conservation laws is to reduce the number of equations that need to be solved. For example, from (3.15) to (3.17) we have

$$B = B_0 + \beta(A - A_0)/\alpha, \quad (3.20)$$

$$C = C_0 + \gamma(A_0 - A)/\alpha, \quad (3.21)$$

$$D = D_0 + \delta(A_0 - A)/\alpha. \quad (3.22)$$

Therefore, once we know  $A$  we will then be able to determine the other three concentrations. From (3.14), the reduced equation for  $A$  takes the form

$$\frac{dA}{dt} = -\alpha k A^\alpha (a + bA)^\beta, \quad (3.23)$$

where  $a = B_0 - bA_0$  and  $b = \beta/\alpha$ . This is still a formidable equation but we only have to deal with one rather than four as was originally stated.

One thing to keep in mind when looking for conservation laws is that they are not unique. For example, you can multiply a conservation law by any nonzero number and the result is another conservation law. The objective is not to find all possible combinations but, rather, a set of independent laws from which all others can be found. It is possible to develop a theory for determining the independent laws and this will be discussed in Sect. 3.4. However, there is a simple test that can be used on most of the examples in this textbook for determining if the laws are independent, and it is given next.

**Independence Test.** A set of conservation laws are independent if each law contains a species that does not appear in the other laws.

So, for example, (3.15)–(3.17) are independent since only one contains  $B$ , only one contains  $C$ , and only one contains  $D$ . Although this test might be self-evident, a formal proof is given in Exercise 3.15.

### 3.2.3 Steady States

In addition to the conservation laws, we will also be interested in the steady-state solutions. To be a steady state the concentration must be constant and it must satisfy the kinetic equations. From (3.23) there are two steady states, one is  $A = 0$  and the second is  $A = -a/b$ . The corresponding steady-state values for the other species in the reaction are determined from (3.20) to (3.22). The one restriction we impose is that the concentrations are nonnegative. Because of this, if  $a/b > 0$ , then the only one physically relevant steady-state solution of (3.23) is  $A = 0$ .

### 3.2.4 Examples

#### Example 1 $A \rightarrow 2C$

The rate of the reaction is  $r = kA$ , and so, the kinetic equations are

$$\begin{aligned}\frac{dA}{dt} &= -kA, \\ \frac{dC}{dt} &= 2kA.\end{aligned}$$

A conservation law is obtained by noting  $\frac{d}{dt}(2A + C) = 0$ . From this it follows that  $C = C_0 + 2A_0 - 2A$ .

From the above kinetic equations, for a steady state it is required that  $A = 0$ . The corresponding steady-state value for  $C$  is determined using the conservation law, which yields  $C = C_0 + 2A_0$ . It is worth noting that the steady state can be obtained directly from the reaction. Namely, if one starts out with  $A_0$  molecules of  $A$ , and each of these is transformed into two molecules of  $C$ , then the reaction will continue until  $A$  is exhausted, so  $A = 0$ , and the amount of  $C$  has increased by  $2A_0$ . ■

#### Example 2 $A + 2B \rightarrow C$

The rate of the reaction is  $r = kAB^2$  and the kinetic equations are

$$\frac{dA}{dt} = -kAB^2, \tag{3.24}$$

$$\frac{dB}{dt} = -2kAB^2, \tag{3.25}$$

$$\frac{dC}{dt} = kAB^2. \tag{3.26}$$

To find the conservation laws, note that (3.24) and (3.26) can be added to yield  $\frac{dA}{dt} + \frac{dC}{dt} = 0$ . It follows that  $A + C = A_0 + C_0$ . Similarly, (3.25) and (3.26) can be combined to yield  $\frac{d}{dt}(B + 2C) = 0$ , from which it follows that  $B + 2C = B_0 + 2C_0$ .



The conclusion is that  $C = C_0 + A_0 - A$  and  $B = B_0 + 2(A - A_0)$ . The resulting reduced equation for  $A$  is therefore

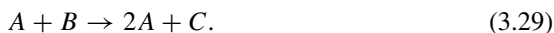
$$\frac{dA}{dt} = -kA(B_0 - 2A_0 + 2A)^2. \quad (3.27)$$

It is possible to use separation of variables to solve this equation, although the solution is determined implicitly (a similar situation will arise in Sect. 3.6.3).

From (3.27), the steady states are  $A = 0$  and  $A = A_0 - \frac{1}{2}B_0$ . For  $A = 0$ , from the conservation laws,  $B = B_0 - 2A_0$  and  $C = C_0 + A_0$ . This is physically relevant only if  $B_0 \geq 2A_0$ . Similarly, for  $A = A_0 - \frac{1}{2}B_0$ , then  $B = 0$  and  $C = C_0 + \frac{1}{2}B_0$ , and this solution requires that  $2A_0 \geq B_0$ . It is also evident from (3.24) and (3.25), assuming  $A$  and  $B$  start out nonzero, that  $A$  and  $B$  decrease monotonically to their steady-state value.

It is possible to determine the steady states directly from the reaction. Namely, the reaction will stop when either  $A$  or  $B$  is exhausted. If it is  $B$ , so  $B = 0$ , then the number of  $A$ 's that are left is  $A_0 - \frac{1}{2}B_0$  and the corresponding value for  $C$  is  $C_0 + \frac{1}{2}B_0$ . A similar deduction can be made when it is  $A$  that is used up. ■

### Example 3



We need to explain what is written here. First, (3.28) is a compact way to write  $A \rightarrow C + D$  and  $C + D \rightarrow A$ . In this case the reactions are said to be reversible. Each gets its own rate constant and we will use  $k_1$  for the first and  $k_{-1}$  for the second. Second, (3.29) is an example of an *autocatalytic reaction* because  $A$  is being used to produce more of itself (i.e., there is more  $A$  at the end of the reaction even though it is one of the reactants). We will use  $k_2$  for its rate constant. The corresponding rates are  $r_1 = k_1A$ ,  $r_{-1} = k_{-1}CD$ , and  $r_2 = k_2AB$ . Now, the Law of Mass Action applies to each reaction and the rates are added to construct the kinetic equation for each species. For example, the kinetic equation for  $A$  is

$$\begin{aligned} \frac{dA}{dt} &= -r_1 + r_{-1} - r_2 + 2r_2 \\ &= -k_1A + k_{-1}CD - k_2AB + 2k_2AB \\ &= -k_1A + k_{-1}CD + k_2AB. \end{aligned}$$

Note that for the reaction in (3.29),  $A$  is treated as both a reactant ( $-r_2$ ) and a product ( $+2r_2$ ) as specified by the reaction. In a similar manner the kinetic equations for the other species are

$$\frac{dB}{dt} = -k_2 AB, \quad (3.30)$$

$$\frac{dC}{dt} = k_1 A - k_{-1} CD + k_2 AB, \quad (3.31)$$

$$\frac{dD}{dt} = k_1 A - k_{-1} CD. \quad (3.32)$$

The useful conservation laws in this case are  $\frac{d}{dt}(A + 2B + C) = 0$  and  $\frac{d}{dt}(A + B + D) = 0$ . From this, we get  $C = A_0 + 2B_0 + C_0 - A - 2B$  and  $D = A_0 + B_0 + D_0 - A - B$ . This enables us to reduce the system to the two equations

$$\begin{aligned} \frac{dA}{dt} &= -k_1 A + k_{-1}(\alpha - A - 2B)(\beta - A - B) + k_2 AB, \\ \frac{dB}{dt} &= -k_2 AB, \end{aligned}$$

where  $\alpha = A_0 + 2B_0 + C_0$  and  $\beta = A_0 + B_0 + D_0$ .

As for the steady states, from (3.30) we have that either  $A = 0$  or  $B = 0$ . For example, if  $A = 0$ , then from (3.31) we have either  $C = 0$  or  $D = 0$ . For the case of  $C = 0$ , from the conservation laws, we get  $B = B_0 + \frac{1}{2}(A_0 + C_0)$  and  $D = D_0 + \frac{1}{2}(A_0 - C_0)$ . One can calculate the other solutions in the same way.

What is interesting is that the reactions paint a slightly different picture about the steady state. The two reactions in (3.28) are no help in finding the steady states as  $A$  simply converts back and forth with  $C$  and  $D$ . The reaction in (3.29), however, will stop when  $B$  is exhausted. In fact,  $B = 0$  is the only apparent species with a steady state. The fact that the reactions give us a different conclusion from what we derived from the differential equations has to do with stability. The differential equations give all mathematically possible steady states irrespective of whether they can be achieved physically. The reactions contain this information by the way they are stated, although they are limited in what they can tell us (e.g., what happens to the other concentrations). What is needed is to introduce the mathematical tools to study stability and this will be done later in the chapter. ■

### 3.2.5 End Notes

It was stated that the coefficient  $k$  in the rate of a reaction is constant. In reality,  $k$  can depend on the conditions under which the reaction takes place. For example, the rate of a chemical reaction depends strongly on the temperature. The most widely used assumption concerning this dependence is the Arrhenius equation, which states

$$k = k_0 e^{-E/RT},$$

where  $k_0$ ,  $E$ , and  $R$  are parameters and  $T$  is temperature measured in Kelvin units. The complication here is that chemical reactions can release, or absorb, heat, and for this reason the temperature can change as the reaction proceeds. It is assumed in this text that the reactions take place in a medium that allows for maintaining a constant temperature.

As a second comment, one might conclude from the physical interpretation of (3.12) that a reaction involving three reactants is rare as it requires three molecules to collide simultaneously. However, it is quite common to find models that contain reactions involving three or more reactants. It is also not unusual to find models with fractional coefficients. This is one of the reasons for introducing the idea of an elementary reaction. These are reactions in which the molecular steps are exactly as stated in the reaction statement. In this case the stoichiometric coefficients equal the number of molecules involved in the reaction. In chemical applications all elementary reactions are either first- or second-order. The fact is, however, that for most reactions the elementary steps are not known. There are multiple reasons for this, but typically it is due to the small concentrations and short life times of the intermediate species, which makes measuring them difficult. Consequently, nonelementary reactions are used and they should be thought of as an approximation of the actual reaction mechanism.

Finally, even though the Law of Mass Action is based on a collection of physically motivated assumptions, the formulation is heuristic. For example, in explaining the dependence of the reaction rate in (3.11) on the species concentrations, we introduced the idea of collision frequency. The fact is that two molecules do not necessarily combine when colliding, and the actual event depends on the collision energy, collision angle, etc. This has led to research into using molecular dynamics to derive the Law of Mass Action from more fundamental principles. This is outside the scope of this textbook and those interested should consult Houston (2006) or Henriksen and Hansen (2008).

### 3.3 Modeling Using the Law of Mass Action

What is of interest is how to translate observations about how species interact into a mathematical model. Two examples are considered, both of which are well-established models arising in biochemistry and epidemiology. The third example illustrates how to determine if a given set of kinetic equations are consistent with mass action.

### 3.3.1 Michaelis-Menten Kinetics

Many chemical and biological systems depend on enzymes to catalyze one or more of their component reactions. Often the exact mechanisms are not well understood and can involve very complicated pathways with multiple enzymes and other catalysts. A relatively simple description of the mechanism is provided by the Michaelis-Menten model, and the reaction steps are shown in Fig. 3.2. The assumption is that an enzyme ( $E$ ) will bind to a substrate ( $S$ ), forming a complex ( $C$ ). The enzyme within the complex will then either modify the substrate and release it as a new species ( $P$ ), or else it will release the substrate unchanged.

According to what is shown in the schematic, the reaction steps are as follows:



Using the Law of Mass Action, the resulting kinetic equations are

$$\frac{dS}{dt} = -k_1SE + k_{-1}C, \quad (3.35)$$

$$\frac{dE}{dt} = -k_1SE + k_{-1}C + k_2C, \quad (3.36)$$

$$\frac{dC}{dt} = k_1SE - k_{-1}C - k_2C, \quad (3.37)$$

$$\frac{dP}{dt} = k_2C. \quad (3.38)$$

For initial conditions, it is assumed that we start with  $S$  and  $E$  and no complex or product. In other words,

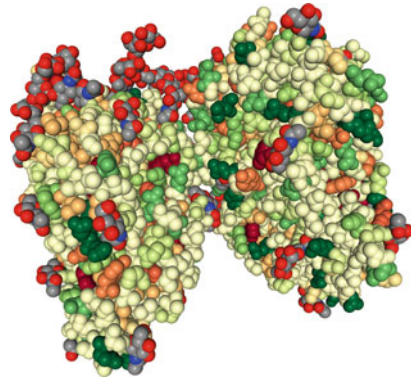
$$S(0) = S_0, E(0) = E_0, C(0) = 0, P(0) = 0. \quad (3.39)$$

Two useful conservation laws for this reaction are  $\frac{d}{dt}(E + C) = 0$  and  $\frac{d}{dt}(S + C + P) = 0$ . Using the stated initial conditions, the conservation laws give us that



**Fig. 3.2** The steps in the Michaelis-Menten mechanism, where an enzyme,  $E$ , assists  $S$  in transforming into  $P$

**Fig. 3.3** The three-dimensional structure of the enzyme invertase (Ramírez-Escudero et al., 2016)



$E = E_0 - C$  and  $P = S_0 - S - C$ . Therefore, we can reduce the reactions to the two equations

$$\frac{dS}{dt} = -k_1 E_0 S + (k_{-1} + k_1) S C, \quad (3.40)$$

$$\frac{dC}{dt} = k_1 E_0 S - (k_2 + k_{-1} + k_1) S C. \quad (3.41)$$

As for the steady states, from (3.38) we get that  $C = 0$ . Using (3.40), we get that  $S = 0$ , and from the two conservation laws we conclude that  $E = E_0$  and  $P = S_0$ .

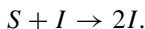
As a historical note, the steps involved in the overall reaction were first written down by Henri (1903) in connection with his study of the hydrolysis of sucrose. The equations were later analyzed by Michaelis and Menten (1913). As it turns out, the hydrolysis of sucrose produces two simpler sugars, glucose and fructose. The enzyme in this case is invertase, also known as  $\beta$ -fructofuranosidase, and the structure of this molecule is shown in Fig. 3.3. Clearly the depiction of the mechanism in Fig. 3.2 is simplistic, but it still provides an effective description of the overall reaction. It is also worth noting that the discovery of this particular reaction represents the beginning of enzyme kinetics as a scientific discipline, and for this reason it has become one of the standard examples in biochemistry courses. It also has commercial applications. The splitting of sucrose into simpler sugars is called inversion, and the mixture produced is called invert sugar. Apparently, according to The Sugar Association, invert sugar is sweeter than regular sugar and this has useful applications in baking and candy making.

### 3.3.2 Disease Modeling

One approach to model how a disease moves through a population is based on separating the population into groups. Physically realistic models usually involve a large number of different groups determined by age, medical history, etc. We will consider a simple case, where the groups consist of the population  $S$  of those

susceptible to the disease, another is the population  $I$  that is ill (and infectious), and the third is the population  $R$  of individuals that have recovered. How, or if, individuals move from group to group is described as follows:

1. If a susceptible individual comes into contact with someone who is ill, they too can become ill. Writing this statement using reactions, we get that



In applying this to a real disease, there are probabilities or uncertainties that are associated with getting sick. This includes the probability of coming into contact with someone who is ill, and also the uncertainty that you will get sick even if you do meet them. These are accounted for with the rate constant for this particular interaction.

2. It is assumed that anyone who is ill will eventually recover, at which point they are immune from the disease. In terms of a reaction, this can be written as



3. It is assumed that a person's immunity can wear off, and they again become susceptible. The resulting reaction is



Combining our results, the kinetic equations for our hypothetical disease are:

$$\frac{dS}{dt} = -k_1 SI + k_3 R, \quad (3.42)$$

$$\frac{dI}{dt} = -k_2 I + k_1 SI, \quad (3.43)$$

$$\frac{dR}{dt} = -k_3 R + k_2 I. \quad (3.44)$$

The one useful conservation law for this reaction system is  $\frac{d}{dt}(S + I + R) = 0$ . This is simply the statement that the total population  $N = S + I + R$  is constant, which is not unexpected since our model does not include deaths or births. One consequence of this conservation law, and the fact that a population cannot be negative, is that  $0 \leq S \leq N$ ,  $0 \leq I \leq N$ , and  $0 \leq R \leq N$ .

An interesting question is whether there is a steady state. From (3.43), for a steady state we need either  $I = 0$  or else  $S = k_2/k_1$ . If  $I = 0$ , then from (3.44),  $R = 0$ , and the conservation law then gives us that  $S = N$ . This is the preferred steady state because, if possible, it means the disease has disappeared from the population. In contrast, the other steady state, which has  $S = k_2/k_1$ ,  $I = k_3 R/k_2$ , and  $R = (N - S)/(1 + k_1/k_2)$ , has a nonzero value for  $I$ , and this gives rise to what is often called an epidemic equilibrium. To be physically relevant, it is required that  $S \leq N$ ,

which means that  $k_2/k_1 \leq N$ . This brings us to one of the more important questions in disease modeling, which is: does the disease disappear from the population or does an epidemic equilibrium occur? How to determine this will be addressed later when considering the stability properties of a steady state.

### 3.3.3 Reverse Mass Action

A question that often arises is, given a set of equations, is it possible to determine if they are consistent with the Law of Mass Action. As an example, consider the predator-prey equations

$$\frac{dR}{dt} = aR - bRW, \quad (3.45)$$

$$\frac{dW}{dt} = -cW + dRW, \quad (3.46)$$

where the coefficients  $a$ ,  $b$ ,  $c$ , and  $d$  are positive constants. Because the rates of the reactions add, we can break this problem down to individual terms that are of similar type. Also, the key to this is to remember that the terms on the right-hand side of the differential equations are determined using the reactants.

$R' = aR$ : The reactant here is  $R$ , and its stoichiometric coefficient is one. So, the reaction must have the form  $R \rightarrow \text{product}(s)$ . This, by itself, will result in the equation  $R' = -k_1 R$ . We need a positive rate, and so  $R$  must also be a product of this reaction. In other words, the reaction has the form  $R \rightarrow \alpha R$ , which gives us the equation  $R' = k_1(\alpha - 1)R$ . For a positive rate it is required that  $\alpha > 1$ . The value of  $\alpha$  is not determinable using the law of mass action.

$R' = -bRW$ ,  $W' = dRW$ : The reactants here are  $R$  and  $W$ , and the stoichiometric coefficients are both one. The resulting reaction must have the form  $R + W \rightarrow \text{product}(s)$ . This yields the equations  $R' = -k_2 RW$  and  $W' = k_2 RW$ . This equation for  $R$  is what we want, but the rate for  $W$  must be positive. So,  $W$  must also be a product, and this means the reaction is  $R + W \rightarrow \beta W$ . The  $R$  equation is unchanged and the one for  $W$  now becomes  $W' = k_2(\beta - 1)RW$ . For positivity, it is required that  $\beta > 1$ .

$W' = -cW$ : The reactant here is  $W$ , and its stoichiometric coefficient is one. The reaction has the form  $W \rightarrow \text{product}$ , and the corresponding equation is  $W' = -k_3 W$ , which is consistent with the original. It makes no difference what the product is other than it does not involve  $R$  or  $W$ . Given the application being considered here, it is appropriate to introduce a new variable  $P$  that represents the number of dead predators. The resulting reaction is then  $W \rightarrow P$ .

We have found that the predator-prey equations can be associated with the following reactions:

$$R \rightarrow \alpha R, \quad (3.47)$$

$$R + W \rightarrow \beta W, \quad (3.48)$$

$$W \rightarrow P. \quad (3.49)$$

Expressing the equations in reaction form provides a different viewpoint on the assumptions that were used to formulate the model. For example, the  $aR$  term in (3.45) comes from the assumption that the number of rabbits increases at a rate proportional to the current population. This is commonly assumed but in looking at (3.47) it is hard to justify, at least for rabbits. For example, without some statement to the contrary, the above model applies to a population of all male rabbits just as well as to one where males and females are evenly distributed. Clearly, it must be assumed that both genders are present for the model to make any sense. Even so, the assumption that a rabbit undergoes mitosis and splits into two rabbits, as implied by (3.47), is a stretch. Another possibility is to redefine  $R$  and assume it represents the population of only the female rabbits. This makes (3.47) somewhat easier to understand, but it raises questions about why the male rabbits do not affect the population of wolves as implied in (3.48). Clearly this is not possible.

The point of the above discussion is that models, by their very nature, are based on assumptions and it is important to have an understanding of what the assumptions are. There is nothing wrong with using a simple model to help develop an understanding of the problem, but it is essential to know what is assumed when this is done.

### 3.4 General Mathematical Formulation

For the general form of the schemes considered here we assume there are  $n$  reactions involving  $m$  distinct species  $X_1, X_2, \dots, X_m$ . The scheme is composed of a set of reactions of the form

$$\sum_{i=1}^m \alpha_{ij} X_i \rightarrow \sum_{i=1}^m \beta_{ij} X_i \quad \text{for } j = 1, 2, \dots, n. \quad (3.50)$$

In this setting a species can appear as just a reactant, so  $\beta_{ij} = 0$ , or as just a product, so  $\alpha_{ij} = 0$ , or both. Also, the stoichiometric coefficients  $\alpha_{ij}, \beta_{ij}$  are assumed to be nonnegative, with at least one of the  $\alpha$ 's and one of the  $\beta$ 's nonzero in each reaction. The assumption that they are nonnegative differs slightly from the earlier requirement that they are positive. This has no effect on the reactions, or the resulting conservation laws, and is done for mathematical convenience.

The reaction rate  $r_j$  for the  $j$ th reaction is



$$r_j = k_j \prod_{i=1}^m X_i^{\alpha_{ij}}, \quad (3.51)$$

where  $k_j$  is the rate constant for the  $j$ th reaction. With this, the kinetic equation for the time evolution of the concentration of  $X_i$  is

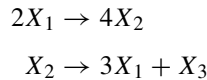
$$\frac{d}{dt}X_i = \sum_{j=1}^n (\beta_{ij} - \alpha_{ij}) r_j. \quad (3.52)$$

This can be written in matrix form as

$$\frac{d}{dt}\mathbf{X} = \mathbf{S}\mathbf{r}, \quad (3.53)$$

where  $\mathbf{X} = (X_1, X_2, \dots, X_m)^T$  is the vector of concentrations and  $\mathbf{r} = (r_1, r_2, \dots, r_n)^T$  is the rate vector. The  $m \times n$  matrix  $\mathbf{S}$  is called the *stoichiometric matrix* and  $S_{ij} = \beta_{ij} - \alpha_{ij}$ . It might help to remember how  $\mathbf{S}$  is formed by noting it is a *species*  $\times$  *reactions* matrix.

*Example*



For these reactions, the stoichiometric matrix is

$$\mathbf{S} = \begin{pmatrix} -2 & 3 \\ 4 & -1 \\ 0 & 1 \end{pmatrix},$$

and the rate vector is

$$\mathbf{r} = \begin{pmatrix} k_1 X_1^2 \\ k_2 X_2 \end{pmatrix}.$$

The resulting matrix problem is

$$\frac{d}{dt} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} -2 & 3 \\ 4 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} k_1 X_1^2 \\ k_2 X_2 \end{pmatrix}.$$

This is the form given in (3.53). ■

A conservation law for the general reaction scheme in (3.50) satisfies

$$\frac{d}{dt}(a_1 X_1 + a_2 X_2 + \cdots + a_m X_m) = 0,$$

where the  $a_j$ 's are constants that will be determined later. Integrating this equation we obtain

$$a_1 X_1 + a_2 X_2 + \cdots + a_m X_m = a_1 X_{10} + a_2 X_{20} + \cdots + a_m X_{m0},$$

where  $X_{j0}$  is the initial concentration of  $X_j$ . It is convenient to express this in vector form, which is

$$\mathbf{a} \cdot \mathbf{X} = \mathbf{a} \cdot \mathbf{X}_0, \quad (3.54)$$

where  $\mathbf{a} = (a_1, \dots, a_m)^T$  and  $\mathbf{X}_0 = (X_{10}, X_{20}, \dots, X_{m0})^T$ .

The vector  $\mathbf{a}$  can be determined by taking the dot product of (3.53) with  $\mathbf{a}$ , to obtain  $\mathbf{a} \cdot \mathbf{X}' = \mathbf{a} \cdot \mathbf{S}\mathbf{r}$ . Given that  $\mathbf{a} \cdot \mathbf{X}' = (\mathbf{a} \cdot \mathbf{X})'$  and  $\mathbf{a} \cdot \mathbf{S}\mathbf{r} = \mathbf{r} \cdot \mathbf{S}^T \mathbf{a}$ , then a conservation law corresponds to  $\mathbf{r} \cdot \mathbf{S}^T \mathbf{a} = 0$ . Therefore, any nonzero vector  $\mathbf{a}$  that satisfies

$$\mathbf{S}^T \mathbf{a} = \mathbf{0} \quad (3.55)$$

yields a conservation law. Written this way, finding the conservation laws has been reduced to a linear algebra problem. It is also important to note that there are conservation laws that do not satisfy (3.55), and how this happens will be explained at the end of this section.

The solutions of (3.55) form a subspace known as the null space of  $\mathbf{S}^T$ . Let  $N(\mathbf{S}^T)$  designate this subspace. If  $N(\mathbf{S}^T)$  contains only the zero vector, so  $\mathbf{a} = \mathbf{0}$  is the only solution of (3.55), then there are no conservation laws coming from (3.55). Assuming there are nonzero solutions, then let  $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k\}$  be a basis for  $N(\mathbf{S}^T)$ . Each basis vector produces a conservation law of the form in (3.54), and it is independent of the laws obtained from the other basis vectors. What this means is that the conservation law obtained using, say,  $\mathbf{a}_1$  cannot be obtained by combining the conservation laws obtained using  $\mathbf{a}_2, \mathbf{a}_3, \dots, \mathbf{a}_k$ . Moreover, because these vectors form a basis, given any conservation law coming from (3.55), we are able to write it in terms of the laws obtained from the basis vectors. These observations are summarized in the following result.

**Null Space Theorem.** *The basis vectors of the null space of  $\mathbf{S}^T$  correspond to independent conservation laws of the system.*

To be specific, if (3.55) has only the zero solution, then no conservation laws are obtained. Otherwise, the resulting conservation laws have the form

$$\mathbf{a}_i \cdot \mathbf{X} = \mathbf{a}_i \cdot \mathbf{X}_0, \quad \text{for } i = 1, 2, \dots, k, \quad (3.56)$$

where  $\mathbf{X}_0 = (X_{10}, X_{20}, \dots, X_{m0})^T$  are the initial concentrations and the vectors  $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k\}$  form a basis of the null space of  $\mathbf{S}^T$ . The benefit of this is that each independent conservation law can be used to eliminate one of the differential equations in (3.53).

In schemes only involving a few reactions it is not necessary to use the Null Space Theorem because you can usually just look at the equations and determine the independent conservation laws. However, for systems containing many equations the above result is very useful as it provides a systematic method for reducing the problem.

*Example (Cont'd)* For the previous example, (3.55) takes the form

$$\begin{pmatrix} -2 & 4 & 0 \\ 3 & -1 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Forming the augmented matrix and row reducing yields the following:

$$\left( \begin{array}{ccc|c} -2 & 4 & 0 & 0 \\ 3 & -1 & 1 & 0 \end{array} \right) \rightarrow \left( \begin{array}{ccc|c} 1 & -2 & 0 & 0 \\ 0 & 5 & 1 & 0 \end{array} \right). \quad (3.57)$$

The solution is  $a_3 = -5a_2$  and  $a_1 = 2a_2$ . Consequently, the null space has dimension one and a basis is  $\mathbf{a}_1 = (2, 1, -5)^T$ . The corresponding conservation law is  $2X_1 + X_2 - 5X_3 = 2X_{10} + X_{20} - 5X_{30}$ . ■

There are certain reaction systems where (3.55) will miss a conservation law. This can happen when the system contains two, or more, reactions with the same reactants and respective stoichiometric coefficients. To illustrate, suppose the system is  $A \rightarrow 3B$  and  $A \rightarrow 2C$ . The kinetic equations in this case are

$$\begin{aligned} \frac{dA}{dt} &= -(k_1 + k_2)A, \\ \frac{dB}{dt} &= 3k_1A, \\ \frac{dC}{dt} &= 2k_2A, \end{aligned}$$

and two independent conservation laws are  $6A + 2B + 3C = \text{constant}$  and  $3k_1A + (k_1 + k_2)B = \text{constant}$ . The latter law comes from the fact that  $r_1 = k_1A$  is a constant multiple of  $r_2 = k_2A$ . Because of this, the requirement that  $\mathbf{r} \cdot \mathbf{S}^T \mathbf{a} = 0$  reduces in this example to  $\mathbf{b} \cdot \mathbf{S}^T \mathbf{a} = 0$ , where  $\mathbf{b} = (k_1, k_2)^T$ . So, a conservation law will result if there is a nonzero vector in the range of  $\mathbf{S}^T$  that is orthogonal to  $\mathbf{b}$ . This vector will depend on the values of  $k_1$  and  $k_2$ , and this means  $\mathbf{a}$  depends on these values.

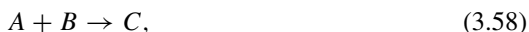
As a final comment, the definition of a conservation law used here is more general than what is often used in the chemical literature. When the subject was in its infancy, to qualify as a conservation law, the coefficients  $a, b, c, \dots, z$  in (3.18) were required to be positive (Horn and Jackson, 1972; Schuster and Hofer, 1991). This is a consequence of the assumption that all chemical reactions are reversible, and this is explained in Exercise 3.13. More recently, the convention in chemistry is that conservation laws come exclusively from the null space of  $\mathbf{S}^T$  (Famili and Pálsson, 2003; Polettini and Esposito, 2014; Reder, 1988). The viewpoint taken here is simply that a conservation law is a linear combination of the concentrations that is time invariant. This is the usual mathematical requirement for a conservation law, and it allows for a greater reduction of the kinetic equations.

## 3.5 Steady States and Stability

As with all time-dependent problems, one of the central questions is what happens to the solution as time increases. Specifically, if a physical system is started out with particular initial conditions, is it possible to determine if the solution will approach a steady state? There are various ways to address this question and we will consider three.

### 3.5.1 Reaction Analysis

It is possible, in some cases, to determine the steady states and their stability properties directly from the reactions. This has been done in several of the examples earlier in the chapter, but as another example consider the reactions



In molecular terms, the reaction  $A + B \rightarrow C$  states that one molecule of  $A$  combines with one molecule of  $B$  to form one  $C$ . The first reaction will continue until  $A$  or  $B$  is used up. However, since  $C$  breaks down into three  $B$ 's, it is not possible to run out of  $B$ . In other words, the first reaction stops when  $A = 0$ . The second reaction continues until  $C$  is gone. From beginning to end, one  $A$  is turned into two  $B$ 's, and the initial  $C$ 's are turned into three  $B$ 's. So, when the reactions run to completion, you end up with  $B = 2A_0 + B_0 + 3C_0$ . This will happen no matter what the initial concentrations, so long as  $B_0$  or  $C_0$  are nonzero. This is the property underlying the idea of a stable steady state. We will define stability shortly, but what is significant is that we have been able to obtain this conclusion without explicitly using the kinetic equations.

### 3.5.2 Geometric Analysis

A second method for analyzing steady states involves a geometric argument using the phase plane. To explain, suppose the kinetic equations are

$$\frac{ds}{d\tau} = -s + (\mu + s)c, \quad (3.60)$$

$$\varepsilon \frac{dc}{d\tau} = s - (\kappa + s)c, \quad (3.61)$$

where  $\varepsilon$ ,  $\mu$ , and  $\kappa$  are positive, with  $\mu < \kappa$ . This example comes from the nondimensional version of the Michaelis-Menten system (see Sect. 3.6.3). It is not necessary to use a nondimensional version, but it does simplify some of the steps involved.

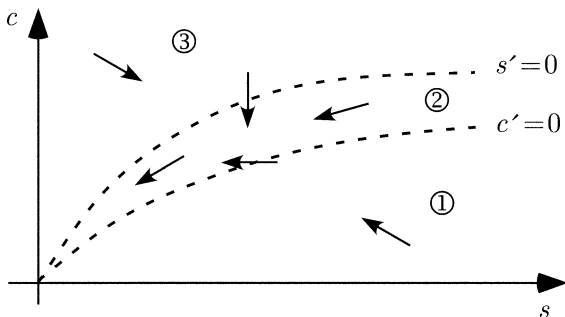
To transform this into the phase plane we combine the above two differential equations to obtain

$$\frac{dc}{ds} = \frac{s - (\kappa + s)c}{\varepsilon(-s + (\mu + s)c)}. \quad (3.62)$$

The idea here is that in the  $c, s$ -plane the solution is a parametric curve, with the time variable  $\tau$  acting as the parameter. With this viewpoint, (3.60) and (3.61) are expressions for the velocities of the respective variables. We will use these equations, along with (3.62), to sketch the solution. Before doing this note that the physically relevant solution satisfies  $0 \leq s$  and  $0 \leq c$ . Limiting our attention to this region, then the situation is sketched in Fig. 3.4.

The first step used to construct Fig. 3.4 was to sketch the two nullclines. The  $s$ -nullcline is the curve where  $s'(\tau) = 0$ , and from (3.60) this is  $c = s/(\mu + s)$ . Similarly, the  $c$ -nullcline is the curve where  $c'(\tau) = 0$ , and from (3.61) this is  $c = s/(\kappa + s)$ . The points where the nullclines intersect are the steady states, and for this problem this is simply  $s = c = 0$ . These two curves separate the quadrant into three regions, designated ①, ②, and ③. Using (3.62) we have the following cases.

**Fig. 3.4** Phase plane and direction fields for the Michaelis-Menten system, in the region  $0 \leq s$  and  $0 \leq c$



*Region ①* In this region  $s' < 0$  and  $c' > 0$ . Consequently,  $\frac{dc}{ds} < 0$  and this means the slope of the solution curve in this region is negative. The slope of the small arrow indicates this in the figure. The arrow points in the direction of motion, and this is determined from the inequalities  $c' > 0$  and  $s' < 0$ . So,  $c$  is increasing and  $s$  is decreasing.

*Region ②* Now  $\frac{dc}{ds} > 0$ , and so the slope of the solution curve in this region is positive. The two small line segments indicate this in the figure. The arrows on these lines are determined by noting from (3.60) that  $c' > 0$  while, from (3.60),  $s' < 0$ . So,  $c$  and  $s$  are decreasing.

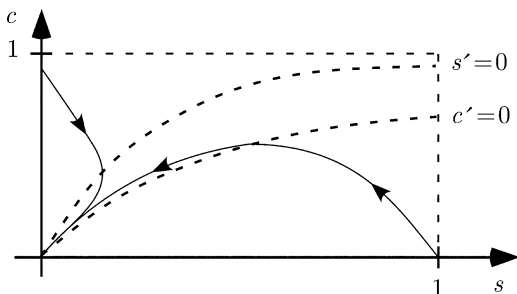
*Region ③* Because  $\frac{dc}{ds} < 0$ , then the slope of the solution curve in this region is negative. The small line segment indicates this in the figure. The arrow on the line is determined by noting from (3.60) that  $c' < 0$  while, from (3.60),  $s' > 0$ . So,  $c$  is decreasing and  $s$  is increasing.

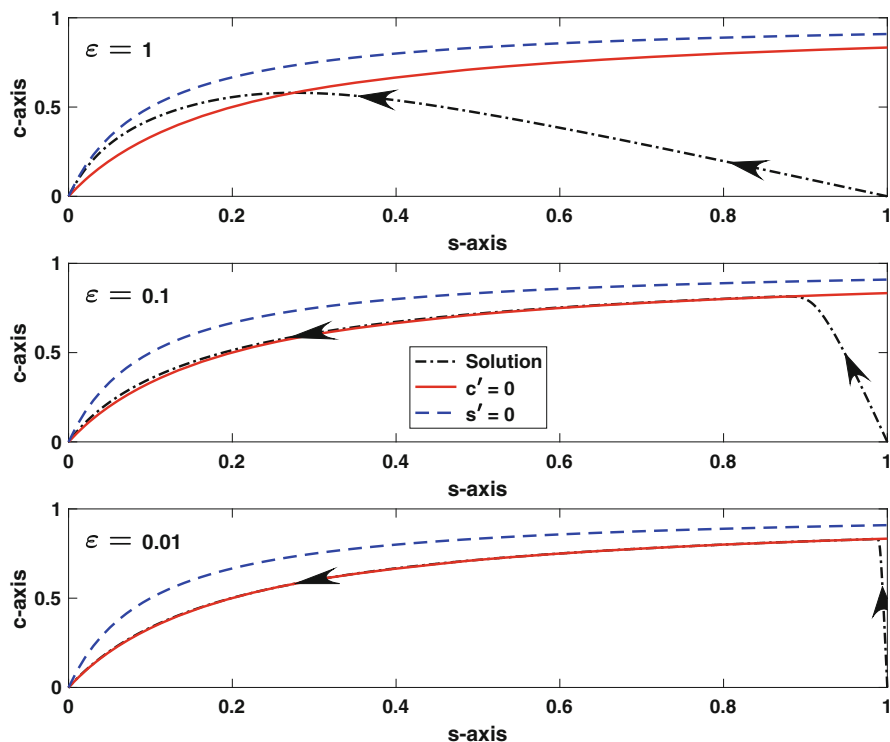
*Nullclines* When  $s' = 0$  the slope of the solution curve, as determined from (3.62), is vertical. When this occurs,  $c$  is decreasing and this explains the arrows. When  $c' = 0$  the slope of the solution curve, as determined from (3.62), is horizontal. When this occurs,  $s$  is decreasing and this explains the arrows.

With this information we are in position to sketch the solution. For the initial condition  $s(0) = 1$  and  $c(0) = 0$ , which means the starting point is in region ① in Fig. 3.4,  $c$  increases and  $s$  decreases. This continues until the  $c$ -nullcline is crossed, after which the solution heads towards  $c = s = 0$ . This is shown in Fig. 3.5. In contrast, if one starts out with  $s = 0$  and  $c \neq 0$ , then  $c$  decreases while  $s$  increases. This continues until the  $s$ -nullcline is crossed, after which the solution converges to  $c = s = 0$ . It would appear that no matter where we start that the same conclusion is reached, and for this reason we conclude that  $s = c = 0$  is a globally asymptotically stable steady state. By global it is understood that the conclusion holds for every initial condition that is in the region  $s \geq 0$  and  $c \geq 0$ .

As a check on the geometric arguments made here, the numerical solution is shown in Fig. 3.6 for various values of  $\varepsilon$ . In each case the two nullclines are also plotted. All three graphs behave as predicted in Fig. 3.5. What is most striking, however, is how the trajectory changes as  $\varepsilon$  decreases. In the case of when  $\varepsilon = 0.01$ ,

**Fig. 3.5** Phase plane and integral curves for the Michaelis-Menten system





**Fig. 3.6** Numerical solution of (3.60) and (3.61) for various values of  $\varepsilon$ , along with the  $s$ - and  $c$ -nullclines. The arrows on the solution curves show the direction of motion. Note the rapid initial rise in  $c$  for small values of  $\varepsilon$

the solution starts out going almost straight up to the  $c$ -nullcline, and then follows that curve into the origin. The resulting solution curve has a passing resemblance to the boundary layer solutions in the previous chapter. It should come as no surprise later, when solving the Michaelis-Menten problem, that a boundary layer approximation is used to find a composite approximation of the solution.

### 3.5.3 Perturbation Analysis

The two methods we have used to study the properties of the steady-state solution have the advantage of not requiring us to solve the differential equations. This makes them very attractive but they are limited in their applicability. For example, using a reaction analysis, if one has even a moderate number of reactions it can be difficult to sort out what species reach a steady state. Similarly, the geometric approach is limited to systems that can be reduced down to two species. We now consider a more analytical method, one capable of resolving a large number of reactions.

Assume that the kinetic equations can be written in vector form as

$$\frac{d}{dt}\mathbf{y} = \mathbf{f}(\mathbf{y}), \quad (3.63)$$

or in component form as

$$\begin{aligned} \frac{d}{dt}y_1 &= f_1(y_1, y_2, \dots, y_n), \\ \frac{d}{dt}y_2 &= f_2(y_1, y_2, \dots, y_n), \\ &\vdots \\ \frac{d}{dt}y_n &= f_n(y_1, y_2, \dots, y_n). \end{aligned}$$

In this case,  $\mathbf{y}_s$  is a steady-state solution if  $\mathbf{y}_s$  is constant and  $\mathbf{f}(\mathbf{y}_s) = \mathbf{0}$ . We say that  $\mathbf{y}_s$  is *stable* if any solution that starts near  $\mathbf{y}_s$  stays near it. If, in addition, initial conditions near  $\mathbf{y}_s$  actually result in the solution converging to  $\mathbf{y}_s$  as  $t \rightarrow \infty$ , then  $\mathbf{y}_s$  is said to be *asymptotically stable*.

To investigate stability we introduce the initial condition

$$\mathbf{y}(0) = \mathbf{y}_s + \delta \mathbf{a}. \quad (3.64)$$

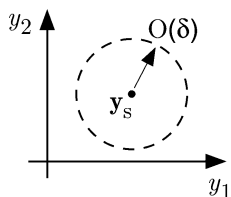
The idea here is that we are starting the solution close to the steady state (see Fig. 3.7), and so we assume  $\delta$  is small. Now, for asymptotic stability it is required that the solution of the resulting initial value problem converges to  $\mathbf{y}_s$  as time increases, irrespective of the particular values for  $\mathbf{a}$ . This will be determined using asymptotics, and the appropriate expansion of the solution for small  $\delta$  is

$$\mathbf{y}(t) \sim \mathbf{y}_s + \delta \bar{\mathbf{y}}(t) + \dots. \quad (3.65)$$

If it is found that

$$\lim_{t \rightarrow \infty} \bar{\mathbf{y}} = \mathbf{0}, \quad (3.66)$$

**Fig. 3.7** As given in (3.64), the initial condition used in the linearized stability analysis is taken to be within a  $O(\delta)$  region around the steady-state solution  $\mathbf{y}_s$





no matter what we pick for  $\mathbf{a}$ , then the steady state is asymptotically stable (to small perturbations). This approach is called a *linear stability analysis* and it will be our standard method for deciding if a steady state is stable.

Before substituting (3.65) into (3.63), note that, using Taylor's theorem, and the chain rule,

$$\begin{aligned} \mathbf{f}(\mathbf{y}_s + \delta \bar{\mathbf{y}} + \cdots) &= \mathbf{f}\Big|_{\delta=0} + \delta \frac{d}{d\delta} \mathbf{f}\Big|_{\delta=0} + \frac{1}{2} \delta^2 \frac{d^2}{d\delta^2} \mathbf{f}\Big|_{\delta=0} + \cdots \\ &= \mathbf{f}(\mathbf{y}_s) + \delta \mathbf{A} \bar{\mathbf{y}} + \cdots, \end{aligned} \quad (3.67)$$

where  $\mathbf{A} = \mathbf{f}'(\mathbf{y}_s)$  is the Jacobian matrix for  $\mathbf{f}$  evaluated at  $\mathbf{y}_s$ . For those who might not remember, the Jacobian for  $\mathbf{f}$  is defined as

$$\mathbf{f}'(\mathbf{y}) \equiv \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \frac{\partial f_1}{\partial y_2} & \cdots & \frac{\partial f_1}{\partial y_n} \\ \frac{\partial f_2}{\partial y_1} & \frac{\partial f_2}{\partial y_2} & \cdots & \frac{\partial f_2}{\partial y_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial y_1} & \frac{\partial f_n}{\partial y_2} & \cdots & \frac{\partial f_n}{\partial y_n} \end{pmatrix}. \quad (3.68)$$

There is not a standard notation for the Jacobian matrix, and other often used choices include  $\mathbf{J}_f$ ,  $D\mathbf{f}$ , and  $\nabla \mathbf{f}$ . The one used here was popularized by Rudin (1976).

Now, since  $\mathbf{f}(\mathbf{y}_s) = \mathbf{0}$ , then the  $O(\delta)$  problem coming from substituting (3.65) into (3.63) and (3.64) is

$$\frac{d}{dt} \bar{\mathbf{y}} = \mathbf{A} \bar{\mathbf{y}}, \quad (3.69)$$

where

$$\bar{\mathbf{y}}(0) = \mathbf{a}. \quad (3.70)$$

The solution is found by assuming that  $\mathbf{y} = \mathbf{x}e^{rt}$ . Substituting this into (3.69), the problem reduces to solving

$$\mathbf{A}\mathbf{x} = r\mathbf{x}. \quad (3.71)$$

This is an eigenvalue problem, where  $r$  is the eigenvalue and  $\mathbf{x}$  is an associated eigenvector. With this, the values for  $r$  are determined by solving the characteristic equation

$$\det(\mathbf{A} - r\mathbf{I}) = 0, \quad (3.72)$$

where  $\mathbf{I}$  is the identity matrix. Given a value of  $r$ , its eigenvectors are then determined by solving  $(\mathbf{A} - r\mathbf{I})\mathbf{x} = \mathbf{0}$ .

We begin with the case of when  $\mathbf{A}$  is not defective. This means that  $\mathbf{A}$  has  $n$  linearly independent eigenvectors  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ , with corresponding eigenvalues  $r_1, r_2, \dots, r_n$ . The  $r_i$  values do not need to be different, but they can be complex-valued. The resulting general solution of (3.69) has the form

$$\bar{\mathbf{y}} = \alpha_1 \mathbf{x}_1 e^{r_1 t} + \alpha_2 \mathbf{x}_2 e^{r_2 t} + \dots + \alpha_n \mathbf{x}_n e^{r_n t}, \quad (3.73)$$

where  $\alpha_1, \alpha_2, \dots, \alpha_n$  are arbitrary constants. The latter are determined from the initial condition (3.70). However, it is not necessary to calculate their values as we are only interested in what happens when  $t \rightarrow \infty$ . With this in mind, note that if  $\text{Re}(r) < 0$ , then

$$\lim_{t \rightarrow \infty} e^{rt} = 0,$$

and if  $\text{Re}(r) > 0$ , then

$$\lim_{t \rightarrow \infty} |e^{rt}| = \infty.$$

Therefore, from (3.73), it follows that  $\bar{\mathbf{y}} \rightarrow \mathbf{0}$  as  $t \rightarrow \infty$  if  $\text{Re}(r_i) < 0, \forall i$ . However, if even one eigenvalue has  $\text{Re}(r_i) > 0$ , then it is possible to find values for  $\mathbf{a}$  so  $\bar{\mathbf{y}}$  is unbounded as  $t \rightarrow \infty$ .

If  $\mathbf{A}$  does not have  $n$  linearly independent eigenvectors, which means that  $\mathbf{A}$  is defective, then the general solution contains  $e^{rt}$  terms as well as those of the form  $t^k e^{rt}$ , where  $k$  is a positive integer (Braun, 1993). Consequently, the conclusion is the same, which is that  $\bar{\mathbf{y}} \rightarrow \mathbf{0}$  as  $t \rightarrow \infty$  if  $\text{Re}(r_i) < 0, \forall i$ . Moreover, if there is an eigenvalue with  $\text{Re}(r_i) > 0$ , then  $\bar{\mathbf{y}}$  can become unbounded as  $t \rightarrow \infty$ .

The remaining question is what happens if one or more of the eigenvalues have  $\text{Re}(r) = 0$ , but the others satisfy  $\text{Re}(r) < 0$ . It is possible to show that if the geometric and algebraic multiplicities of a  $\text{Re}(r) = 0$  eigenvalue are not equal, then  $\mathbf{y}_s$  is unstable. Aside from this, the linear Taylor series approximation of  $\mathbf{f}$  used here is not capable of determining the stability of the steady state in this particular case. As simple examples illustrating the various possible stability outcomes, you should work out the following cases:  $y' = -y^3$ ,  $y' = y^3$ , and  $y' = 0$ .

The discussion in the previous paragraphs gives rise to the next result.

**Linearized Stability Theorem.** *Based on a linear stability analysis, the steady-state  $\mathbf{y}_s$  is asymptotically stable if all of the eigenvalues of  $\mathbf{A}$  satisfy  $\text{Re}(r) < 0$ , and it is unstable if even one eigenvalue has  $\text{Re}(r) > 0$ .*

An interesting aspect of this theorem is that it does not actually require determining the eigenvalues. To explain, the characteristic equation (3.72), when multiplied out, has the form

$$r^n + a_1 r^{n-1} + \dots + a_{n-1} r + a_n = 0. \quad (3.74)$$

What we want to know is, do the roots of this polynomial satisfy  $\text{Re}(r) < 0$ ? Determining the properties of the roots of a polynomial has been studied for centuries, and several tests have been developed that are useful for answering our stability question. This includes Descartes' rule of signs, Sturm sequences, and the Hurwitz conditions. A few of the more useful tests are included in the following theorem (see Exercise 3.28).

**Special Cases Stability Theorem.**

- i) If  $n = 2$ , then a steady state is asymptotically stable if  $\text{tr}(\mathbf{A}) < 0$  and  $\det(\mathbf{A}) > 0$ . It is unstable if either of these inequalities is reversed.
- ii) If  $n = 3$ , then a steady state is asymptotically stable if  $a_1 > 0$ ,  $a_3 > 0$ , and  $a_1 a_2 > a_3$ . It is unstable if any of these inequalities is reversed.
- iii) The steady state is unstable if the number of changes of sign between consecutive nonzero coefficients in (3.74) is odd.

To illustrate the condition given in statement (iii), there is one sign change for  $r^4 + 2r^3 - 1 = 0$ , one sign change for  $r^4 - 2r^3 - 1 = 0$ , and three sign changes for  $r^4 - r^3 + 9r^2 + 5r - 1 = 0$ . In each case, according to statement (i), the associated steady state is unstable.

*Example 1* We begin with the Michaelis-Menten system derived in Sect. 3.6. It was found that there is one steady state, which is  $(S, C, E, P) = (0, 0, E_0, S_0)$ . To determine if this is stable, it is not necessary to use the original set of four kinetic equations (3.40)–(3.41). Instead, it is possible, and much easier, to use the reduced set of equations obtained using the conservation laws. These are given in (3.40) and (3.41), and they are

$$\frac{dS}{dt} = -k_1 E_0 S + (k_{-1} + k_1 S)C, \quad (3.75)$$

$$\frac{dC}{dt} = k_1 E_0 S - (k_2 + k_{-1} + k_1 S)C. \quad (3.76)$$

In this example,  $\mathbf{y} = (S, C)^T$ ,

$$\mathbf{f} = \begin{pmatrix} -k_1 E_0 S + (k_{-1} + k_1 S)C \\ k_1 E_0 S - (k_2 + k_{-1} + k_1 S)C \end{pmatrix},$$

and

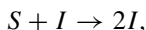
$$\mathbf{f}' = \begin{pmatrix} \frac{\partial f_1}{\partial S} & \frac{\partial f_1}{\partial C} \\ \frac{\partial f_2}{\partial S} & \frac{\partial f_2}{\partial C} \end{pmatrix} = \begin{pmatrix} -k_1 E_0 + k_1 C & k_{-1} + k_1 S \\ k_1 E_0 - k_1 C & -(k_2 + k_{-1} + k_1 S) \end{pmatrix}. \quad (3.77)$$

Setting  $S' = C' = 0$  in (3.75) and (3.76) one finds that the only steady state is  $S_s = C_s = 0$ . Substituting this into (3.77) yields

$$\mathbf{A} = \mathbf{f}'(\mathbf{y}_s) = \begin{pmatrix} -k_1 E_0 & k_{-1} \\ k_1 E_0 & -(k_2 + k_{-1}) \end{pmatrix}.$$

Since  $\text{tr}(\mathbf{A}) = -k_1 E_0 - (k_2 + k_{-1}) < 0$  and  $\det(\mathbf{A}) = k_1 k_2 E_0 > 0$ , then from the above special cases theorem, the steady state  $(S_s, C_s) = (0, 0)$  is asymptotically stable. ■

*Example 2* The reactions for the disease model derived in Sect. 3.3.2 are



and the resulting kinetic equations are given in (3.42)–(3.44). It was determined that there are two steady states, one with  $I = 0$  and a second with  $S = k_2/k_1$ . As in the previous example, the stability question can be answered using the reduced set of kinetic equations. Since the conservation law is  $N = S + I + R$ , then the reduced set of kinetic equations for this example is

$$\frac{dS}{dt} = -k_1 S I + k_3 (N - I - S), \quad (3.78)$$

$$\frac{dI}{dt} = -k_2 I + k_1 S I. \quad (3.79)$$

Consequently,

$$\mathbf{f}' = \begin{pmatrix} \frac{\partial f_1}{\partial S} & \frac{\partial f_1}{\partial I} \\ \frac{\partial f_2}{\partial S} & \frac{\partial f_2}{\partial I} \end{pmatrix} = \begin{pmatrix} -k_1 I - k_3 & -k_1 S - k_3 \\ k_1 I & -k_2 + k_1 S \end{pmatrix}.$$

For the steady state  $(S, I, R) = (N, 0, 0)$ ,

$$\mathbf{A} = \begin{pmatrix} -k_3 & -k_1 N - k_3 \\ 0 & -k_2 + k_1 N \end{pmatrix}.$$

With this,  $\text{tr}(\mathbf{A}) = -k_3 - k_2 + k_1 N$  and  $\det(\mathbf{A}) = k_3(k_2 - k_1 N)$ . For stability we need  $\det(\mathbf{A}) > 0$ , and so it is required that  $N < k_2/k_1$ . We also need  $\text{tr}(\mathbf{A}) < 0$ , and this means we need  $N < (k_2 + k_3)/k_1$ . This holds if  $N < k_2/k_1$ , and so the steady state is asymptotically stable if  $N < k_2/k_1$  and it is unstable if  $N > k_2/k_1$ . ■

*Example 3* As a third example we consider the system

$$\frac{du}{dt} = -v, \quad (3.80)$$

$$\varepsilon \frac{dv}{dt} = u + \lambda(v - v^3/3). \quad (3.81)$$

It is assumed  $\varepsilon$  is positive and  $\lambda$  is nonzero. This is a special case of what is known as the van der Pol equation. Setting  $u' = v' = 0$ , the steady state is found to be  $(u_s, v_s) = (0, 0)$ . With this, and carrying out the required differentiations, it is found that

$$\mathbf{A} = \begin{pmatrix} 0 & -1 \\ \frac{1}{\varepsilon} & \frac{\lambda}{\varepsilon} \end{pmatrix}.$$

Since  $\text{tr}(\mathbf{A}) = \lambda/\varepsilon$  and  $\det(\mathbf{A}) = 1/\varepsilon$ , then from the above theorem, the steady state is asymptotically stable if  $\lambda < 0$  and is unstable if  $\lambda > 0$ .

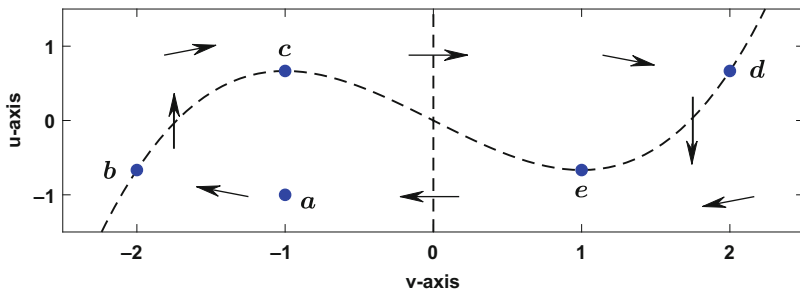
Up to this point, this example appears to be similar to the previous one. However, there is an important difference in how the steady state goes unstable. In the last example, both eigenvalues are real-valued, and the steady state  $(S, I, R) = (N, 0, 0)$  goes unstable because one of the eigenvalues of  $\mathbf{A}$  switches from  $r < 0$  to  $r > 0$ . In the current example, when  $\lambda$  is close to zero, the eigenvalues are complex-valued. To explain why this is important, the eigenvalues for  $\mathbf{A}$  are, for  $\lambda$  close to zero, the two eigenvalues are

$$r_{\pm} = \frac{1}{2\varepsilon} \left( \lambda \pm i\sqrt{4\varepsilon - \lambda^2} \right). \quad (3.82)$$

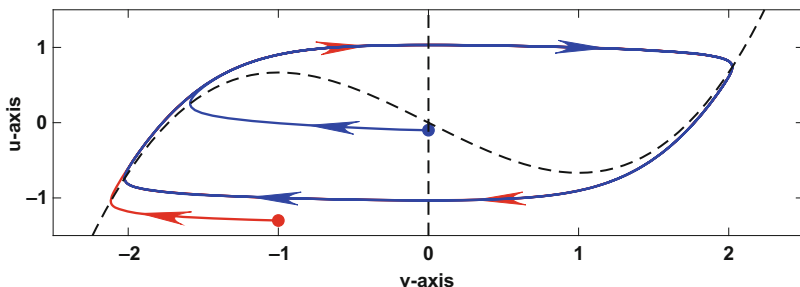
The steady state goes unstable, as  $\lambda$  goes from negative to positive, because both  $r_+$  and  $r_-$  move from the left half-plane, where  $\text{Re}(r) < 0$ , into the right half-plane, where  $\text{Re}(r) > 0$ , as  $\lambda$  passes through zero. Moreover, at  $\lambda = 0$ ,  $\frac{d}{d\lambda} \text{Re}(r) \neq 0$ . Together, these two properties can be used to prove that as  $\lambda$  goes from negative to positive that a stable periodic solution appears for  $\lambda > 0$ . Why this happens is explained below. In the vernacular of dynamical systems,  $\lambda = 0$  is an example of a Hopf bifurcation point, and the periodic solution is a *limit cycle*.

A sketch of the basic properties of the solution in the phase plane is given in Fig. 3.8. The  $u$ -nullcline is the vertical line  $v = 0$ , while the  $v$ -nullcline is the cubic  $u = -\lambda(v - v^3/3)$ . Both are shown in the figure with dashed curves. The small lines indicate the slope as determined from  $\frac{dv}{du}$ , with the arrowheads showing the direction of motion.

Suppose the initial condition corresponds to the point  $a$  in Fig. 3.8. According to the directional arrow in this region, the solution will start out by moving upwards and to the left. It will cross the dashed cubic curve, near the point  $b$ , after which it will move upwards and to the right. It will pass over the point  $c$ , and once it



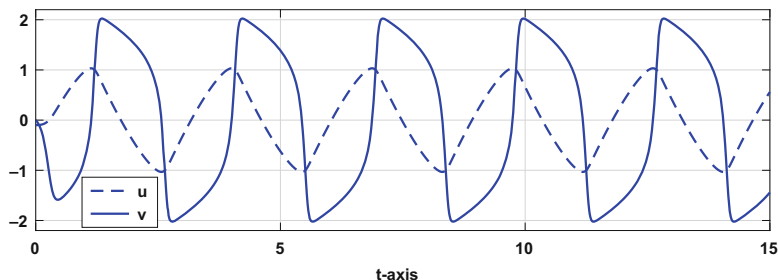
**Fig. 3.8** Phase plane and direction fields for Example 3, in the case of when  $\lambda > 0$



**Fig. 3.9** Numerical solution of (3.80) and (3.81) using different starting points. In the calculations,  $\varepsilon = 0.1$  and  $\lambda = 1$ . Also, the dashed curves are the two nullclines

crosses the vertical dashed line, it will start to move downward and to the right. It will eventually cross the dashed cubic, near the point  $d$ , after which the solution will move downwards and to the left. It will pass under the point  $e$ , and once it crosses the vertical dashed line, it will again start to move upward and to the left. This encircling of the unstable steady state at the origin will repeat over and over again. Moreover, after one or more complete circuits, the solution will approach a closed curve, which is known as a *limit cycle*.

To reinforce the conclusion of the previous paragraph, the numerical solution of the problem is given in Fig. 3.9. Two different starting points are used, one inside and the other outside the limit cycle. In both cases the solution converges to the limit cycle. To illustrate the periodic nature of the solution once it starts following the limit cycle, in Fig. 3.10 the functions  $u(t)$  and  $v(t)$  are shown. These are the same functions that form the blue trajectory in Fig. 3.9. Note that the relatively steep rise in  $v$  seen in Fig. 3.10 corresponds to the rapid movement from  $c$  to  $d$  in Fig. 3.8, while the steep drop comes from moving from  $e$  to  $b$ . ■



**Fig. 3.10** The functions  $u(t)$  and  $v(t)$  corresponding to the blue trajectory shown in Fig. 3.9

### 3.6 Solving the Michaelis-Menten Problem

With the Law of Mass Action it is relatively easy to write down the kinetic equations. We have also developed some useful tools for determining the overall behavior of the solution. We now consider the ultimate question, which is, how do you find the solution? Given that the kinetic equations are usually nonlinear, answering this requires some finesse. There are different ways to proceed and, as in many problems, the choice depends on one's interests and background. We will consider three, one using numerical methods, one based on the rates of the reactions, and then one using perturbation expansions.

For completeness, it is worth restating the problem we are going to try to solve. As shown in Sect. 3.3, it is

$$\frac{dS}{dt} = -k_1 E_0 S + (k_{-1} + k_1 S)C, \quad (3.83)$$

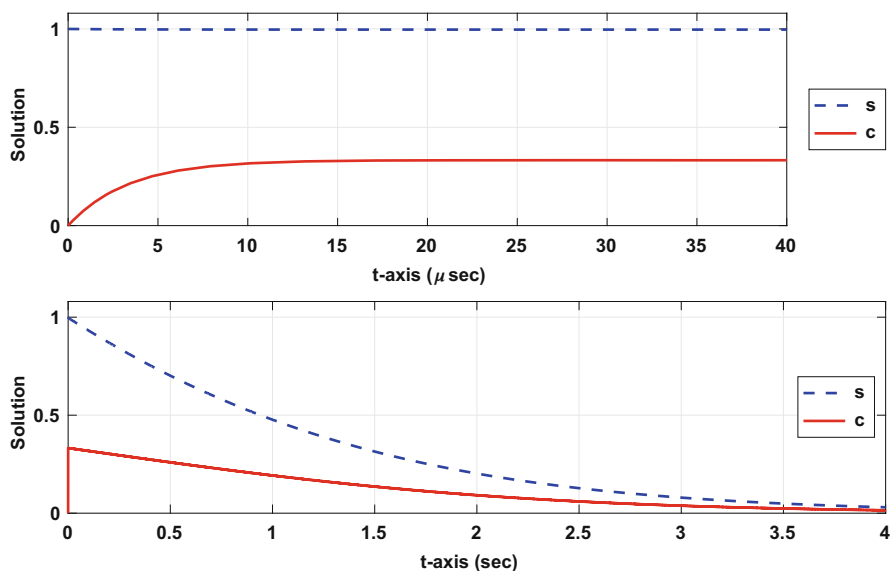
$$\frac{dC}{dt} = k_1 E_0 S - (k_2 + k_{-1} + k_1 S)C, \quad (3.84)$$

where  $S(0) = S_0$  and  $C(0) = 0$ . The other two species are determined using the conservation laws  $E = E_0 - C$  and  $P = S_0 - S - C$ .

#### 3.6.1 Numerical Solution

Solving the problem numerically is straightforward, and one only has to decide on what parameter values to use. We will use those that come from a model of the transport of P-glycoprotein (Agnani et al. 2011). They found that  $k_1 = 10^8 \text{ M}^{-1} \text{ s}^{-1}$ ,  $k_{-1} = 2 \times 10^5 \text{ s}^{-1}$ ,  $k_2 = 200 \text{ s}^{-1}$ ,  $E_0 = 10^{-5} \text{ M}$ , and  $S_0 = 100E_0$ . The resulting solution curves  $s = S/S_0$  and  $c = C/E_0$  are shown in Fig. 3.11.

One of the interesting aspects of this experiment is that the amount of  $E$  is small in comparison to the initial concentration of  $S$ . This is typical because enzymes



**Fig. 3.11** Numerical solution of (3.83) and (3.84) using parameter values for the transport of P-glycoprotein (Agnani et al. 2011). Shown are  $s = S/S_0$  and  $c = C/E_0$ . The upper plot shows the solution curves during the first 40  $\mu\text{s}$  of the experiment, while the lower plot shows the solution curves over the first 4 s

are usually very efficient catalysts, and this is why they are usually present in relatively small concentrations. A second interesting feature of the solution is that  $c$  changes relatively quickly, within the first few microseconds of the experiment. For example, as shown in the top plot of Fig. 3.11,  $c$  starts at zero but increases up to approximately 0.33. In the lower plot, it looks like  $c$  simply jumps from  $c = 0$  up to about  $c = 0.33$ . This is the type of behavior that would be expected if there were a boundary layer, although in this problem, because we are looking at the time variable, it is more appropriate to refer to this as an initial layer. There are a couple of ways to take advantage of the initial jump in  $c$ , and these are explained below.

### 3.6.2 Quasi-Steady-State Approximation

The rapid change in  $C$  was evident to the experimentalists who first studied this reaction scheme. The physical reasoning usually given is that, assuming the concentration of  $S$  is not too small, the enzyme is so efficient that whenever an  $E$  becomes free that it immediately attaches itself to an  $S$  to form another complex  $C$ . The implication is that the concentration of  $C$  changes so quickly in response to the values of the other species that it immediately satisfies its steady-state equation. This is the basis of what is known as a *quasi-steady-state assumption (QSSA)*, an



idea first proposed by Briggs and Haldane (1928). The argument made is that the equations can be replaced with

$$\frac{dS}{dt} = -k_1 E_0 S + (k_{-1} + k_1 S)C, \quad (3.85)$$

$$0 = k_1 E_0 S - (k_2 + k_{-1} + k_1 S)C. \quad (3.86)$$

Solving the last equation for  $C$  yields

$$C = \frac{k_1 E_0 S}{k_2 + k_{-1} + k_1 S}. \quad (3.87)$$

To put this approximation in the context of the transport of P-glycoprotein, as given in Fig. 3.11, it is seen that the concentration of  $S$  changes on a time scale of seconds. This is why  $S$  appears to be constant in the upper plot. In comparison,  $C$  changes on a much faster time scale, measured in microseconds. This means  $C$  adjusts to the value of  $S$  so quickly, moving to what it assumes is the steady state, that its value is determined by the formula in (3.87). The exception to this statement is what happens at the very beginning of the experiment, where  $C$  must undergo a jump to be able to satisfy (3.87).

Mathematically one should be a bit skeptical with this approximation. For one, the formula in (3.87) does not satisfy the given initial condition for  $C$ . For another, because  $S$  is time-dependent,  $C$  in (3.87) clearly depends on  $t$ , and this does not appear to be consistent with the assumption used to derive this result. These questions will be addressed once the perturbation solution is derived. For the moment we will assume all is well and in this case the equation for  $S$ , given in (3.85), can be written as

$$\frac{dS}{dt} = -\frac{v_M S}{K_M + S}. \quad (3.88)$$

In this equation  $v_M = k_2 E_0$ , and  $K_M$  is the Michaelis constant given as

$$K_M = \frac{k_{-1} + k_2}{k_1}. \quad (3.89)$$

Experimentalists use (3.88) to determine  $v_M = k_1 E_0$  and  $K_M$  by measuring  $S'$  at  $t = 0$ . The specifics of how this is done are explored in Exercise 3.29. As it turns out, experimental studies that determine all three rate constants are not common. There are technical, and mathematical, challenges in determining these constants, and an indication of what is involved can be found in Tran et al. (2005) and Agnani et al. (2011).

### 3.6.3 Perturbation Approach

The QSSA is one of the standard methods used by biophysicists, and mathematicians, to reduce a reaction scheme. As pointed out in the derivation, there are several mathematical questions concerning the consistency of the assumptions and for this reason we now consider a perturbation approximation. The underlying hypothesis in the analysis is that it takes very little enzyme to convert  $S$  to  $P$ . In other words, it is assumed that  $E_0$  is much smaller than  $S_0$ .

#### 3.6.3.1 Nondimensionalization

The first step in analyzing the solution is to nondimensionalize the problem. For  $S$  we use its initial condition  $S(0) = S_0$  and set  $S = S_0 s$ , where  $s$  is the nondimensional version of  $S$ . The initial condition for  $C$  is not much help here, but the conservation law  $E + C = E_0$  is because it indicates that the concentration of  $C$  can range from zero up to  $E_0$ . Based on this observation, we take  $C = E_0 c$ , where  $c$  is the nondimensional version of  $C$ . It is not clear what to use for the time variable, and so we simply set  $t = t_c \tau$ , where  $\tau$  is the nondimensionalized time variable. Introducing these into (3.83) and (3.84) and cleaning things up a bit, produces the equations

$$\frac{1}{t_c k_1 E_0} \frac{ds}{d\tau} = -s + (\mu + s)c, \quad (3.90)$$

$$\frac{1}{t_c k_1 S_0} \frac{dc}{d\tau} = s - (\kappa + s)c, \quad (3.91)$$

where  $\mu = k_{-1}/(k_1 S_0)$  and  $\kappa = (k_{-1} + k_2)/(k_1 S_0)$ . We are left with two dimensionless groups that involve  $t_c$ , and one of them will be set to one to determine  $t_c$ . The conventional choice is to use the group in (3.90), and with this  $t_c = 1/(k_1 E_0)$ . In this case the Michaelis-Menten problem becomes

$$\frac{ds}{d\tau} = -s + (\mu + s)c, \quad (3.92)$$

$$\varepsilon \frac{dc}{d\tau} = s - (\kappa + s)c, \quad (3.93)$$

where

$$s(0) = 1, \quad c(0) = 0, \quad (3.94)$$

and

$$\varepsilon = \frac{E_0}{S_0}. \quad (3.95)$$

We are assuming  $\varepsilon$  is small, and because it is multiplying the highest derivative in (3.93) we have a singular perturbation problem. For this reason it should be no surprise that we will find that the function  $c$  has a layer, at  $t = 0$ , where it undergoes a rapid transition. A consequence of this is that  $c$  quickly reaches what is called a quasi-steady state and, for all intents and purposes,  $s - (\kappa + s)c = 0$ . Models containing fast dynamics, which in this case is the  $c$  equation, and slow dynamics, the  $s$  equation, are common in applications. In this way, the Michaelis-Menten system serves as a prototype enzymatic reaction as well as a prototype fast-slow dynamical system. Exactly how this happens, and what to do about the initial condition for  $c$ , will be derived using a perturbation argument.

The equations in (3.92) and (3.93) are a relatively straightforward perturbation problem. We will concentrate on deriving the first term in the approximation and the first step is the outer expansion.

### 3.6.3.2 Outer Expansion

The appropriate expansions are  $s \sim s_0 + \varepsilon s_1 + \dots$  and  $c \sim c_0 + \varepsilon c_1 + \dots$ . Inserting these into (3.92) and (3.93) and collecting the order  $O(1)$  terms we get

$$\frac{ds_0}{d\tau} = -s_0 + (\mu + s_0)c_0, \quad (3.96)$$

$$0 = s_0 - (\kappa + s_0)c_0. \quad (3.97)$$

Solving (3.97) for  $c_0$  and substituting the result into (3.96) gives us

$$\frac{ds_0}{d\tau} = -\frac{\lambda s_0}{\kappa + s_0}, \quad (3.98)$$

where  $\lambda = k_2/(k_1 S_0)$ . Separating variables, and integrating, leads us to the following solution

$$\kappa \ln(s_0) + s_0 = -\lambda \tau + A, \quad (3.99)$$

$$c_0 = \frac{s_0}{\kappa + s_0}, \quad (3.100)$$

where  $A$  is a constant of integration that will be determined later when the layers are matched. The implicit nature of the solution in (3.99) is typical when solving nonlinear differential equations. It is possible to simplify this expression using what is known as the Lambert W function, however (3.99) is sufficient for what we have in mind.

### 3.6.3.3 Inner Expansion

The initial layer coordinate is  $\bar{\tau} = \tau/\varepsilon$  and in this region we will designate the solutions as  $\bar{s}$  and  $\bar{c}$ . The problem is therefore

$$\frac{d\bar{s}}{d\bar{\tau}} = \varepsilon(-\bar{s} + (\mu + \bar{s})\bar{c}), \quad (3.101)$$

$$\frac{d\bar{c}}{d\bar{\tau}} = \bar{s} - (\kappa + \bar{s})\bar{c}, \quad (3.102)$$

where

$$\bar{s}(0) = 1, \quad \bar{c}(0) = 0. \quad (3.103)$$

The appropriate expansions in this case are  $\bar{s} \sim \bar{s}_0 + \varepsilon\bar{s}_1 + \dots$  and  $\bar{c} \sim \bar{c}_0 + \varepsilon\bar{c}_1 + \dots$ . Inserting these into (3.101) and (3.102) and collecting the first-order terms we get

$$\begin{aligned} \frac{d\bar{s}_0}{d\bar{\tau}} &= 0, \\ \frac{d\bar{c}_0}{d\bar{\tau}} &= \bar{s}_0 - (\kappa + \bar{s}_0)\bar{c}_0, \end{aligned}$$

where

$$\bar{s}_0 = 1, \quad \bar{c}_0(0) = 0.$$

Solving these equations gives us

$$\bar{s}_0 = 1, \quad (3.104)$$

$$\bar{c}_0 = \frac{1}{1 + \kappa} \left( 1 - e^{-(1+\kappa)\bar{\tau}} \right). \quad (3.105)$$

The solution for  $\bar{s}$  indicates that, at least to first-order, it does not have an initial layer structure and is constant in this region. The variable  $\bar{c}$  on the other hand does depend on the initial layer coordinate and this is consistent with our earlier numerical experiments.

### 3.6.3.4 Matching and Composite Expansion

The idea underlying matching is the same as used in the previous chapter, namely the solution coming out of the inner layer is the same as the solution going into the inner layer. Mathematically, the requirements are

$$\lim_{\bar{\tau} \rightarrow \infty} \bar{s}_0 = \lim_{\tau \rightarrow 0} s_0, \quad (3.106)$$

$$\lim_{\bar{\tau} \rightarrow \infty} \bar{c}_0 = \lim_{\tau \rightarrow 0} c_0. \quad (3.107)$$

From (3.104) and (3.106) we conclude that  $s_0(0) = 1$ . It is not hard to show that in this case (3.107) is satisfied, and we have that  $A = 1$  in (3.99).

The only remaining task is to construct a composite expansion. This is done by adding the inner and outer solutions and then subtracting their common part. In other words,  $s \sim s_0(\tau) + \bar{s}_0(\bar{\tau}) - s_0(0)$  and  $c \sim c_0(\tau) + \bar{c}_0(\bar{\tau}) - c_0(0)$ . The conclusion is that

$$s \sim s_0(\tau), \quad (3.108)$$

$$c \sim \frac{s_0}{\kappa + s_0} - \frac{1}{1 + \kappa} e^{-(1+\kappa)\tau/\varepsilon}, \quad (3.109)$$

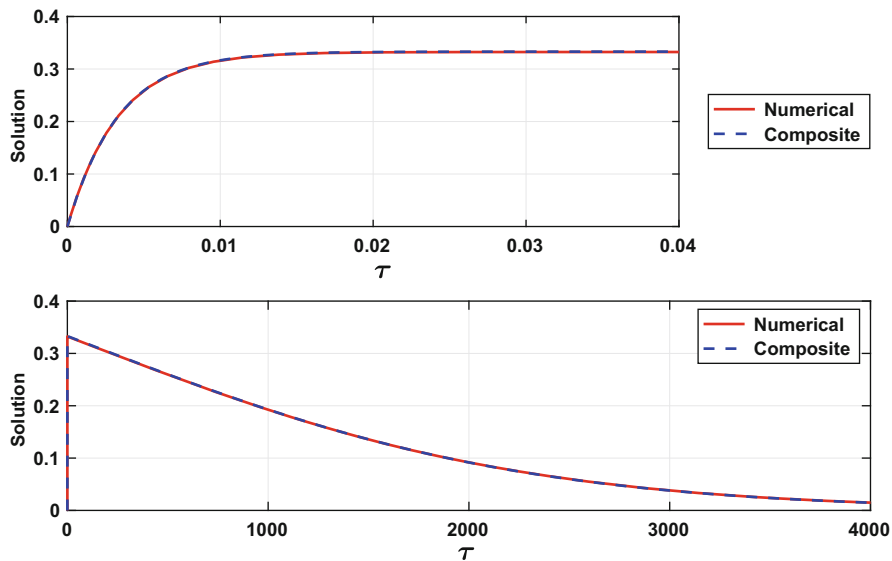
where  $s_0$  is found by solving

$$\kappa \ln(s_0) + s_0 = -\lambda\tau + 1. \quad (3.110)$$

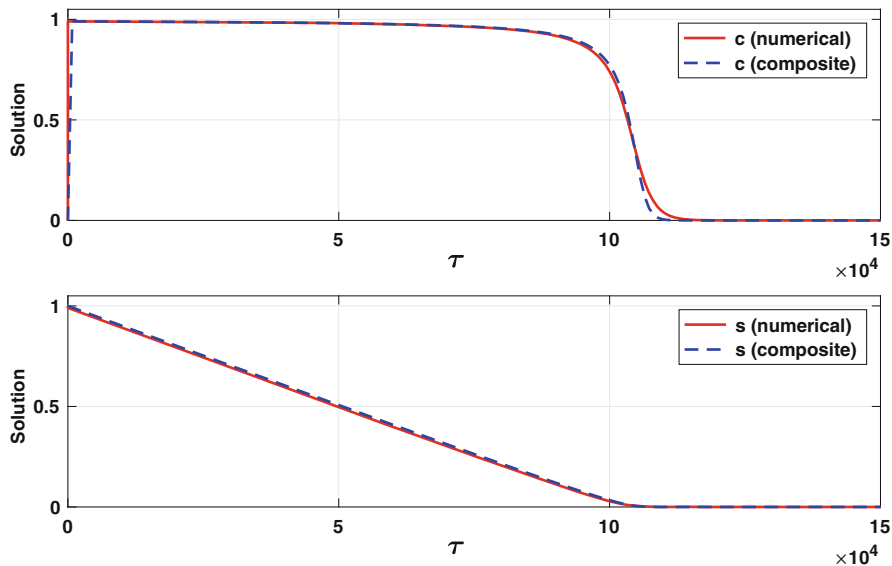
### 3.6.3.5 Analysis of Solution

Although a simple closed form expression for  $s_0$  is not available it is still possible to describe its basic behavior. First, from (3.98) we know it is monotone decreasing. For small values of  $\tau$ , we have from (3.110) that  $s_0 \approx s_0(0) + s'_0(0)\tau = 1 - \lambda\tau/(1 + \kappa)$ . In other words, it starts off by decreasing linearly with slope  $-\lambda/(1 + \kappa)$ . For large  $\tau$  values we have that  $\kappa \ln(s_0) \approx -\lambda\tau$ , and so,  $s_0 \approx e^{-\lambda\tau/\kappa}$ . Therefore,  $s_0$  decays exponentially to zero for large  $\tau$ , with a relaxation time constant  $\kappa/\lambda$ . Based on this information it is an easy matter to sketch the function.

It is also not difficult to use numerical methods, such as Newton's method, to solve the  $s_0$  equation. The resulting composite approximation for  $c(\tau)$  is shown in Fig. 3.12 using the parameter values from the model of the transport of P-glycoprotein (Agnani et al. 2011). Also given is the solution obtained from solving the original equations (3.83) and (3.84) numerically. The agreement is so good that it is difficult to distinguish between the numerical and composite solutions. This agreement is not limited to the P-glycoprotein values, and to demonstrate this another comparison is given in Fig. 3.13 with  $k_1 = 2 \times 10^{10} \text{ M}^{-1}\text{s}^{-1}$  and the other values the same as before. This particular choice was made as it also demonstrates an interesting behavior in the solution of the Michaelis-Menten equations. Namely, it appears that there is a transition layer in the solution as  $k_1$  increases, located in the general vicinity of  $\tau = 1.03 \times 10^5$  (this is when the numerical solution gives  $c = 1/2$ ). With the composite approximations it is easy to explain this behavior. Because  $\kappa$  is very small, we have from (3.109) that  $c \approx 1$  outside the initial layer until  $s$  drops down to the value of  $\kappa$ . In the previous paragraph we know that the



**Fig. 3.12** The solution for  $c(\tau)$  obtained from solving the equations numerically, and from the composite approximation in (3.109) and (3.108). The curves are so close it is difficult to distinguish between them. The parameter values are the same as used for Fig. 3.11



**Fig. 3.13** The solution of (3.92) and (3.93) obtained from solving the equations numerically, and from the composite approximation in (3.109). The parameter values are the same as used for Fig. 3.11 except  $k_1 = 2 \times 10^{10} \text{ M}^{-1} \text{ s}^{-1}$

linear decrease in  $s$  can be approximated as  $s \approx 1 - \lambda\tau$ . Setting this equal to  $\kappa$  and solving we find that  $\tau \approx 1/\lambda$ . With the given parameter values this gives us  $\tau \approx 10^5$ , which is what is seen in Fig. 3.13.

### 3.6.3.6 Connection with QSSA

It is informative to return to the assumptions underlying the quasi-steady-state assumption (3.85) and (3.86). In the outer region, in dimensional coordinates, the equations (3.96) and (3.97) become

$$\begin{aligned}\frac{dS}{dt} &= -k_1 E_0 S + (k_{-1} + k_1 S)C, \\ 0 &= k_1 E_0 S - (k_2 + k_{-1} + k_1 S)C.\end{aligned}$$

These are identical to those used in the quasi-steady-state assumption. In other words, QSSA is effectively equivalent to using an outer approximation of the solution. The actual justification for this type of reduction is the perturbation analysis carried out earlier.

Another observation concerns the resulting equation for  $S$ , which takes the form

$$\frac{dS}{dt} = -\frac{k_1 k_2 E_0 S}{k_2 + k_{-1} + k_1 S}.$$

What is of interest here is that this equation contains a rational function of the variables rather than the power functions expected from the Law of Mass Action. The reason for mentioning this is that one finds models where the equations are rational functions and it is questionable whether they are derivable from mass action. This example demonstrates it is possible although determining this in general can be difficult. This idea will be explored in more depth in Sect. 3.8.

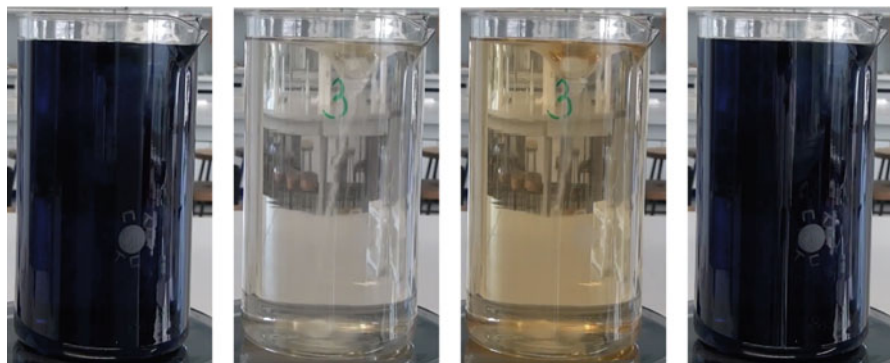
The composite approximation also suggests a possible modification of the analysis. The exponential dependence in (3.109) indicates that the approximation in the inner region holds not just for small  $\varepsilon$ , but also when  $\kappa$  is large. In contrast, we saw in Fig. 3.13 that small values of  $\kappa$  can lead to the appearance of what looks to be a transition layer. To investigate these possibilities it is necessary to modify the nondimensionalization or the expansions used for the solution. This type of post-analysis of the expansion to gain insight into perhaps a better approximation is not uncommon, particularly for difficult problems where it is not clear at the beginning what scales should be used. An analysis of this type related to the Michaelis-Menten equations can be found in Segel and Slemrod (1989).

## 3.7 Oscillators

The Michaelis-Menten reactions resulted in the solution converging, with little fanfare, to a steady-state solution. For other kinetic systems an often occurring solution is one that is oscillatory. This includes those like the predator prey problem that have a periodic solution, with the period depending on the initial conditions. Another type involves a limit cycle, an example of which was considered in Sect. 3.5.3. It is the latter that is of interest here, and how one can arise in a chemical system.

The idea of a chemical oscillator was not accepted easily, and the first paper reporting such a system generated more articles devoted to proving it wrong than trying to understand how it works. This is also true for the most well-known oscillator, the Belousov-Zhabotinskii (BZ) reaction. This was discovered by B. P. Belousov when studying the Krebs cycle. He found that a solution of citric acid in water, with acidified bromate as the oxidant and yellow ceric ions as the catalyst, alternated in color, from yellow to clear, approximately every minute and did this for an hour. His suggestion that this was a form of chemical oscillator was not accepted at the time because it was believed that oscillations in closed homogeneous systems were impossible because that would imply that the reaction did not go smoothly to a thermodynamic equilibrium. About ten years later, A. M. Zhabotinskii expanded on Belousov's research and the work was presented at the 1968 Symposium on Biological and Biochemical Oscillators in Prague. Not unexpectedly, their system has become known as the BZ reaction. Since their early work many other chemical oscillators have been found, and one is shown in Fig. 3.14. In this particular experiment, the system alternates between three states: clear, blue, and amber. As with many such oscillators, the exact mechanism is unknown.

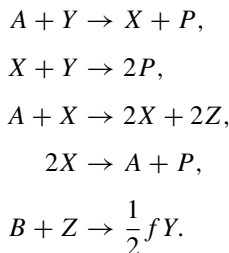
The most widely accepted model for the BZ reaction is due to Field, Körös, and Noyes (Field and Noyes 1974; Field et al. 1972). Their original formulation



**Fig. 3.14** Color changes in a chemical oscillator. These are frames from a video recording of the experiment, the time of the respective frame, from left to right, is 6.3 s, 17.3 s, 19.2 s, and 21.2 s (Thoisoi2, 2014)



contained eleven reactions involving 12 chemical species. It is possible to reduce this system to five reactions, and they are



The chemicals involved here are bromous acid (X), bromide (Y), cerium-4 (Z), bromate (A), an organic species (B), and a product P. The two reactions that stand out are the third because it is autocatalytic, and the last because it involves a rather unusual stoichiometric coefficient. As will be seen below, the parameter  $f$  plays an important role in producing the oscillations in the solution. This reduced model is often called the Oregonator, due to the location of Field and Noyes when they first derived it.

There are two additional simplifying assumptions made in the Oregonator reduction. Namely, in the experiment the concentrations of  $A$  and  $B$  are so large in comparison to the other chemicals that it is assumed they are constant during the reaction. With this the Law of Mass Action produces the following kinetic equations:

$$\begin{aligned}
 \frac{dX}{dt} &= k_1AY - k_2XY + k_3AX - 2k_4X^2, \\
 \frac{dY}{dt} &= -k_1AY - k_2XY + \frac{1}{2}fk_5BZ, \\
 \frac{dZ}{dt} &= 2k_3AX - k_5BZ.
 \end{aligned}$$

To nondimensionalize the problem we take  $X = X_c x$ ,  $Y = Y_c y$ ,  $Z = Z_c z$ , and  $t = t_c \tau$ , where  $X_c = k_3A/(2k_4)$ ,  $Y_c = k_3A/k_2$ ,  $Z_c = (k_3A)^2/(k_4k_5B)$ , and  $t_c = 1/(k_5B)$ . In this case the above equations become

$$\varepsilon x' = \alpha y - xy + x(1 - x), \quad (3.111)$$

$$\delta y' = -\alpha y - xy + fz, \quad (3.112)$$

$$z' = x - z, \quad (3.113)$$

where  $\varepsilon = 4 \times 10^{-2}$ ,  $\alpha = 8 \times 10^{-4}$ , and  $\delta = 4 \times 10^{-4}$ . We will take advantage of the small values of these three parameters in constructing an approximation of the solution.

The objective here is to understand how the species in the reactions are able to produce sustained oscillations over a long period of time. Although we have the tools to handle all three equations, we have learned something from the Michaelis-Menten reaction that enables us to simplify the situation a bit. The very small value of  $\delta$ , which multiplies  $y'$ , means that this particular equation reaches a quasi-steady state very quickly compared to the other two equations. Using this observation we have that  $y = fz/(\alpha + x)$ . With this the equations reduce to

$$\varepsilon x' = x(1 - x) + \frac{\alpha - x}{\alpha + x} fz, \quad (3.114)$$

$$z' = x - z. \quad (3.115)$$

It is this system of equations that we will analyze. In this sense we will be constructing an approximation that is the outer solution to (3.111)–(3.113).

The question arises as to why the QSSA is not also applied to (3.114), which would reduce the entire system down to a single equation. This is an example that illustrates that some care is needed when using the QSSA. As we will see shortly, even though  $x$  tries to reach a quasi-steady state, there are values for the parameter  $f$  for which the equation does not have the capability to reach a steady state. The solution has to repeatedly reposition itself, trying to maintain a quasi-steady state, and this gives rise to a pronounced nonsteady behavior in the solution.

### 3.7.1 Stability

The first step is to determine the steady states. Setting  $x' = 0$  and  $z' = 0$  one finds two solutions,  $(x_s, z_s) = (0, 0)$  and  $(x_s, z_s) = (\bar{x}, \bar{x})$ , where

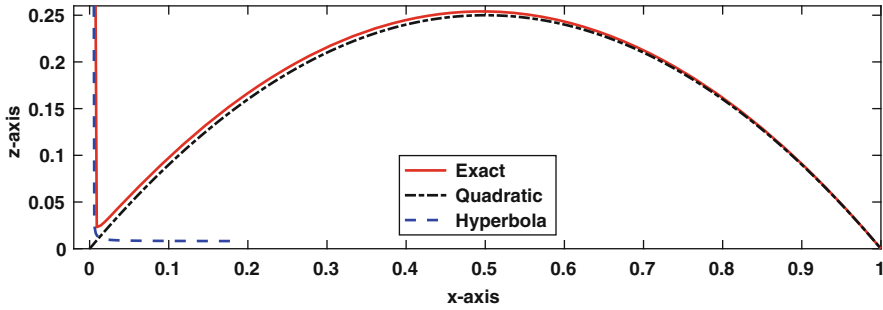
$$\bar{x} = \frac{1}{2} \left[ -\kappa + \sqrt{\kappa^2 + 4\alpha(1 + f)} \right], \quad (3.116)$$

for  $\kappa = \alpha + f - 1$ .

To determine the stability properties of the steady states we will use the geometric argument. The most difficult step for this is to sketch the  $x$ -nullcline, which corresponds to the curve

$$z = -x(1 - x) \frac{\alpha + x}{(\alpha - x)f}. \quad (3.117)$$

A rough sketch can be made by making use of the fact that  $\alpha$  is very small, while  $x$  ranges over the interval  $0 \leq x < \infty$ . Except when  $x$  is near  $\alpha$ , we can use the approximation  $\alpha \pm x \approx \pm x$ . With this (3.117) reduces to



**Fig. 3.15** The  $x$ -nullcline (3.117) is shown along with its quadratic approximation, (3.118), and its hyperbolic approximation, (3.119). For this example,  $f = 1$  and  $\alpha = 4 \times 10^{-3}$

$$z \approx \frac{1}{f}x(1-x). \quad (3.118)$$

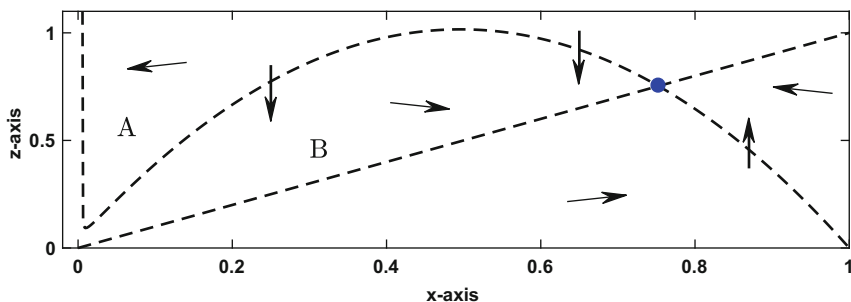
This is simply a quadratic as shown in Fig. 3.15. For  $x$  near  $\alpha$ , then (3.117) reduces to

$$z \approx -\frac{2\alpha x}{(\alpha-x)f}. \quad (3.119)$$

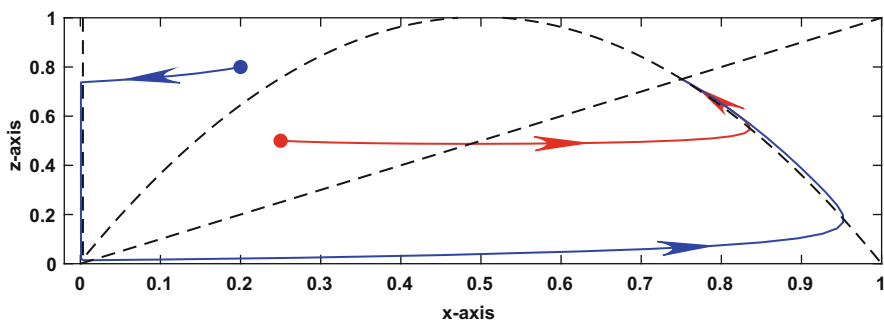
This hyperbola is also shown in Fig. 3.15. With this it is possible to sketch the  $x$ -nullcline. Namely, for  $x$  near  $\alpha$  the curve is given in (3.119), and everywhere else it is given in (3.118). A comparison between these approximations and the exact curve is given in Fig. 3.15.

The above approximations for the  $x$ -nullcline make it easy to estimate where the critical points are located. For example, the local minimum is near where the two approximation curves intersect. Equating (3.118) and (3.119), we have that  $x \approx 3\alpha$ . Similarly, the local maximum comes from the quadratic (3.119), and this is therefore located at  $x \approx \frac{1}{2}$ . Both of these values are rough estimates, and more accurate approximations will be derived shortly. One last point to make here is that the value of  $\alpha$  used in Fig. 3.15 is larger than the value for the BZ system. This was done to make the characteristics of the curve more apparent in the plot because when using the actual value the three curves are so close that it is not possible to distinguish among them graphically.

To use the geometric argument the value of  $f$  needs to be specified. We will consider two cases, and the first is  $f = \frac{1}{4}$ . The resulting phase plane diagram is given in Fig. 3.16. The two nullclines are shown, as are the direction fields. The trajectory of the solution depends on the location of the starting point relative to the  $x$ -nullcline. So, if you start in region A, then you will move quickly to the left, follow the  $x$ -nullcline down to the local minimum, and then move very quickly to the right branch of the  $x$ -nullcline and follow it up to the steady state. Starting in region B you will move quickly to the right to reach the  $x$ -nullcline. To reinforce these conclusions, the numerical solution of the problem is given in Fig. 3.17.



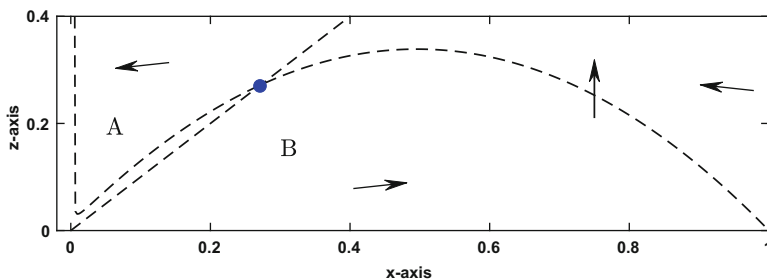
**Fig. 3.16** Phase plane and direction fields for (3.114) and (3.115) when  $f = \frac{1}{4}$



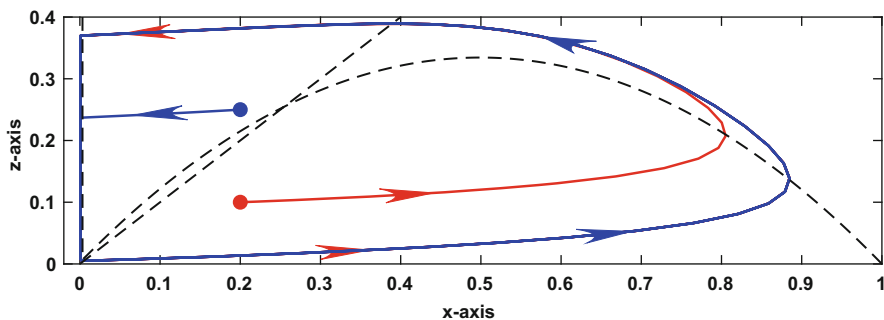
**Fig. 3.17** Numerical solution of (3.114) and (3.115) in the case of  $f = \frac{1}{4}$  for two different initial conditions

The convergence of the solution to the steady state occurs when  $f = \frac{1}{4}$  because the nullclines intersect at a point that the solution is able to reach. As an example of when this does not happen, let  $f = \frac{3}{4}$ . The phase plane, and direction fields, are shown in Fig. 3.18. With a starting point in regions A or B, the solution curve will be very similar to what happened for the  $f = \frac{1}{4}$  case. However, as the solution curve moves up the right branch of the  $x$ -nullcline it will now come close to reaching the local maximum for the nullcline. When this happens, the solution will then move very quickly to the far left branch of the  $x$ -nullcline. Once there the whole process repeats, and what results is a limit cycle. To reinforce this conclusion, the numerical solution of the problem is given in Fig. 3.19.

One of the more apparent differences between the phase planes in Figs. 3.16 and 3.18 is where the two nullclines intersect. This is important as this location determines the stability of the steady state. If the value of  $f$  is such that the nullclines intersect between the local maximum and minimum points of the  $x$ -nullcline, as in Fig. 3.18, then the resulting steady state is unstable. To determine when this occurs, note that the two nullclines intersect, at a nonzero value, when the following holds:



**Fig. 3.18** Phase plane and direction fields for (3.114) and (3.115) when  $f = \frac{3}{4}$



**Fig. 3.19** Numerical solution of (3.114) and (3.115) in the case of  $f = \frac{3}{4}$  for two different initial conditions

$$-(1-x) \frac{\alpha + x}{(\alpha - x)f} = 1. \quad (3.120)$$

As shown in Exercise 3.23, for small  $\alpha$ , the local minimum is at  $x \sim (1 + \sqrt{2})\alpha$ . Consequently the two nullclines intersect at the local minimum when  $f \sim 1 + \sqrt{2}$ . Using a similar analysis one finds that they intersect at the local maximum when  $f \sim \frac{1}{2}$ . Therefore, our conclusion is that the nonzero steady state is unstable if  $\frac{1}{2} < f < 1 + \sqrt{2}$ , and for these values of  $f$  the solution forms a limit cycle similar to the one shown in Fig. 3.19.

### 3.8 Modeling with the QSSA

The stoichiometric coefficients of most reactions are determined experimentally. As an example, suppose it is known that a product  $P$  is produced when two chemicals,  $A$  and  $B$ , are mixed together. The usual empirical-based assumption is that the rate

of formation has the form  $r = kA^\alpha B^\beta$ , and the exponents and rate constant are determined by curve fitting this expression to the experimental measurements of  $r$ . Inevitably, such curve fits produce fractional values for  $\alpha$  and  $\beta$ . This is partly due to the inherent error in the procedure, both experimental as well as numerical. It is also due to the fact that the  $r$  used in the curve fit is really an assumption about the overall rate of the reaction and not a statement about the actual sequence of molecular events through which  $A$  and  $B$  combine to form  $P$ . For example,  $\text{H}_2$  and  $\text{Br}_2$  combine to form hydrogen bromide,  $\text{HBr}$ . Curve fitting the rate function  $r = k\text{H}_2^\alpha \text{Br}_2^\beta$  to the data it is found that  $\alpha = 1$  and  $\beta = 1/2$ . Clearly, this cannot be associated with an elementary reaction, and this begs the question as to just how such an exponent can occur (one possible answer is considered in Exercise 3.27).

To address this question, consider the rather simple looking reaction



This gives rise to the kinetic equation

$$\frac{dA}{dt} = -kAX^2. \quad (3.122)$$

There are equations for the other species but for what we have in mind the above equation is enough. From a physical standpoint this type of reaction is highly unlikely, and the reason is that it requires one  $A$  and two  $X$ 's to come together simultaneously to form the product  $P$ . In the real world, reactions are either unary or binary. The reaction in (3.121) is tertiary because it involves three reactant molecules.

Our objective is to find *elementary reactions* that can give rise to a nonelementary reaction as in (3.121). By elementary we mean the reactions only involve one or two reactant molecules. To do this, the working assumption will be that one or more of the elementary reactions are rapid enough that the QSSA can be employed. It is understood that the expansion into elementary reactions is not necessarily unique and there are possibly multiple ways to do this. Rather, we are interested in whether it can be done at all.

As a first attempt to expand (3.121) into elementary reactions, it is not unreasonable to assume that a molecule of  $A$  combines with  $X$  to form an intermediate complex  $Z$ , and then  $Z$  combines with another  $X$  to form  $P$ . In reaction form this becomes



with the result that

$$\frac{dA}{dt} = -k_1 AX, \quad (3.125)$$

$$\frac{dZ}{dt} = k_1 AX - k_2 XZ. \quad (3.126)$$

The assumption is that the intermediate species  $Z$  plays a role similar to the intermediate species  $C$  in the Michaelis-Menten reaction. As with  $C$ , we will assume  $Z$  reaches a quasi-steady state very quickly and, as in (3.97), this translates into the condition that  $k_1 AX - k_2 XZ = 0$ . Looking at this result a few seconds it is clear that it is not going to help us rewrite (3.125) so it resembles (3.122).

What we are missing is the reversible reaction present in Michaelis-Menten, and so, the reaction scheme is modified to become



The kinetic equations in this case are

$$\frac{dA}{dt} = -k_1 AX + k_{-1} Z, \quad (3.129)$$

$$\frac{dZ}{dt} = k_1 AX - k_{-1} Z - k_2 XZ. \quad (3.130)$$

Imposing a quasi-steady-state assumption on  $Z$  yields

$$Z = \frac{k_1 AX}{k_{-1} + k_2 X}. \quad (3.131)$$

Substituting this into (3.129) gives us that

$$\frac{dA}{dt} = -\frac{k_1 k_2}{k_{-1} + k_2 X} AX^2. \quad (3.132)$$

Making the “*extreme parameter value*” assumption that  $k_2 X \ll k_{-1}$ , then we obtain the kinetic equation in (3.122).

Clearly several assumptions went into this derivation and one should go through a more careful analysis using scaling and perturbation methods to delineate what exactly needs to be assumed. However, the fact that a nonphysical reaction can be explained using elementary reactions has been established for this example.

### 3.9 Epilogue

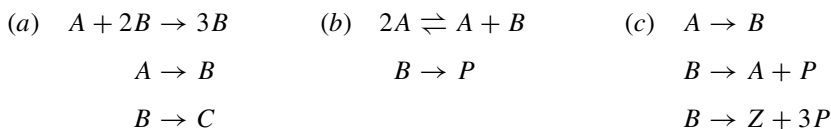
This chapter has introduced some of the foundational tools for deriving, and then analyzing, a mathematical model for multiple interacting species. The subject has a long and rich history, and numerous references for the specific topics that were covered were provided as the material was developed. For more general interest, for those interested in learning more about the theoretical aspects of the subject, you might consult Strogatz (2014) or Hale and Kocak (1996). Those interested in the computational challenges in solving these equations should consider Griffiths and Higham (2010) or Butcher (2016).

What was not discussed, and which usually ends up being the hardest problem to solve, is how to determine the rate constants. A typical model will involve multiple reactions, and, if you are lucky, some of the rate constants will be known. Finding the others requires obtaining the relevant experimental data, and then using one or more data fitting methods to find the  $k$ 's. Being the real world, this almost never works as expected, and you end up using what might be called “creative insights” to finally determine the rate constants. For more about the data fitting methods, you might consult Holmes (2016).

## Exercises

### Section 3.2

**3.1** For the given reactions, determine the kinetic equations and the independent conservation law(s). Also, assuming  $A(0)$  is nonzero, but all other species start out at zero, determine the steady state.



**3.2** This problem considers various aspects of chemical reactions.

- (a) What is the simplest reaction that has rate  $r = kAB^3$ , and conservation law  $A - 3B - 4C = \text{constant}$ ?
- (b) What is the simplest reaction that has rate  $r = kA^2B$ , and conservation law  $A + B = \text{constant}$ ?
- (c) Suppose that the rate constants for the two reactions  $\alpha A + \beta B \rightarrow \text{products}$  and  $\gamma A + \delta C \rightarrow \text{products}$  have the same dimensions. What relationship must hold between the stoichiometric coefficients of the two reactions?



(d) Suppose the kinetic equations include the equation

$$\frac{dA}{dt} = k_1 A + k_2 B + k_3 C.$$

Explain why every steady state for the system must have  $A = B = C = 0$ .

(e) Give an example of a reaction involving a species  $A$  which results in  $\lim_{t \rightarrow \infty} A = \infty$  (i.e., the concentration of  $A$  becomes unbounded).

**3.3** The surface of a solid can act as a catalyst for certain gases. In the Eley-Rideal mechanism it is assumed that a species  $A$  in the gas attaches to a site on the surface, forming a complex  $C$  on the surface. The assumed reaction is  $A + S \rightleftharpoons C$ . Another species  $B$  in the gas then reacts with a surface complex to release a product  $P$  into the gas. The reaction is  $B + C \rightarrow P$ .

- Write down the kinetic equation for each species.
- Find three independent conservation laws and reduce the kinetic equations down to two, for  $S$  and  $C$ .

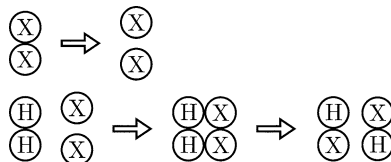
### Section 3.3

**3.4** A well-studied family of reactions involves molecular hydrogen  $H_2$  combining with a diatomic halogen  $X_2$  to produce two molecules of  $HX$ . For example, one can have  $X$  be fluorine (F) or iodine (I). A sequence of steps that lead to the combination of  $H_2$  with  $X_2$  to produce two  $HX$ 's is shown in Fig. 3.20.

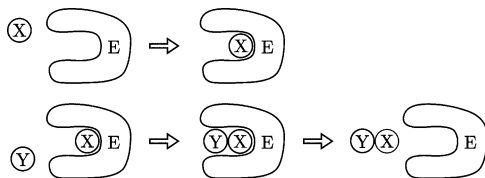
- Write down three reactions corresponding to these steps. Assume the steps are irreversible.
- Derive the rate equations for the five species involved in this sequence of steps.
- Find two independent conservation laws using your equations from part (b). Provide a reason why the laws are independent.
- There is one steady state. Find it using your answers from parts (a) and (c). Make sure to explain your reasoning.

**3.5** Some enzymes work by sequentially binding molecules, and an example is shown in Fig. 3.21. The idea here is that the enzyme  $E$  has a location that has a high

**Fig. 3.20** Figure for Exercise 3.4



**Fig. 3.21** Figure for Exercise 3.5



affinity for binding  $X$ , and the resulting molecule then binds  $Y$ . The last step is the dissociation of the new product molecule  $YX$  from  $E$ .

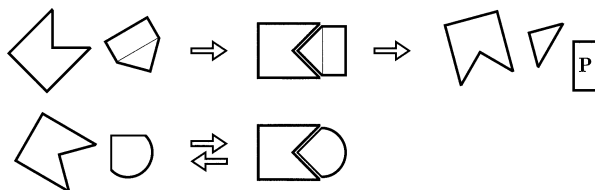
- Write down three reactions corresponding to these steps. Assume the steps are irreversible.
- Derive the rate equations for the six species involved in this sequence of steps.
- Find three independent conservation laws using your equations from part (b). Provide a reason why the laws are independent.
- Using your reactions in part (a) determine the steady state(s). Assume that only  $X$ ,  $Y$ , and  $E$  have nonzero concentrations at the start. Make sure to explain your reasoning.

**3.6** Inhibition occurs when something interferes with, or slows down, a process. The reactions in Fig. 3.22 illustrate how the production of a product  $P$  can be inhibited, and it is an example of what is known as competitive inhibition.

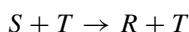
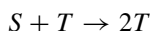
- Label each species and then write down the four reactions corresponding to these steps. Note that  $P$  is already labeled.
- Explain why it is an example of inhibition.
- Derive the rate equations for the seven species involved in this sequence of steps.
- Find four independent conservation laws using your equations from part (c). Provide a reason why the laws are independent.
- Suppose the three species on the far left in Fig. 3.22 start out with nonzero concentrations, and the others start out at zero. From the reactions, initial conditions, and the conservation laws, determine which of the seven species, if any, will approach a steady-state value, and what are the respective values?
- Suppose little, if any, inhibition is observed when the reactions begin, but inhibition becomes increasingly more pronounced as the reactions proceed. Explain how to replace the reversible reaction in Fig. 3.22 with another reversible reaction that would produce this effect. The entire reaction system should only involve six species.

**3.7** The approach used to model disease propagation is often adapted, and modified, to study other things that involve populations. To illustrate this, this exercise explores modeling how a joke moves through a population. In this version of the model there are three groups:  $S$  is the population that either has not heard the joke, or does not remember it,  $T$  is the population of those who know the joke and they will tell it to others, and  $R$  is the population who know the joke but will not tell it to

**Fig. 3.22** Figure for Exercise 3.6



others (they are not good joke tellers or they don't think it's all that funny). A set of reactions involving these three groups is:



- As in Sect. 3.3.2, provide an explanation of what assumption(s) are being made to obtain each reaction.
- Write down the kinetic equation for each species.
- Not all interactions appear in the set of reactions. For example,  $S + R$  and  $R + T$  are not included. What are the products for each of these interactions? Why is it not necessary to include these reactions in the model?
- Find a conservation law for the system.
- There are two steady states, what are they? One of them, to be physically achievable, puts a requirement on the rate constants. Make sure to state what the requirement is.

**3.8** The standard model for population growth is  $P' = kP$ . This leads to exponential growth, which is not sustainable. It is more realistic to have the rate of growth slowdown as the population increases. In fact, if the population is very large, the population should decrease instead of increase. This leads to the assumption that  $k = r_0(1 - P/K)$ , which yields the equation  $P' = r_0P - r_1P^2$ , where  $r_1 = r_0/K$  is a constant. This is known as the logistic equation. The question is whether it is possible to derive it using the Law of Mass Action, using reactions that are plausibly consistent with the application.

- Show that the logistic equation is obtainable using two reactions involving  $P$ . Explain why at least one of the reactions requires an unrealistic assumption about the individuals making up the population.
- Introduce a new species  $F$  that represents the food, or something similar, for the species  $P$ . Using a single reaction involving  $F$  and  $P$ , and a conservation law, derive the logistic equation.

**3.9** The equations below come from applying the Law of Mass Action to two reactions.

$$X' = aX + bYZ,$$

$$Y' = cX + bYZ,$$

$$Z' = cX + dYZ.$$

- (a) Find the two reactions and determine how the coefficients  $a$ ,  $b$ ,  $c$ , and  $d$  are related to each other, if at all. Assume  $a$ ,  $b$ ,  $c$ , and  $d$  are nonzero, but they can be positive or negative.
- (b) Find two independent conservation laws for these reactions.

**3.10** The equations below come from applying the Law of Mass Action to two reactions.

$$X' = aXY,$$

$$Y' = bYZ + cZ,$$

$$Z' = dYZ + eZ.$$

- (a) Find the two reactions and determine how the coefficients  $a$ ,  $b$ ,  $c$ , and  $d$  are related to each other, if at all. Assume  $a$ ,  $b$ ,  $c$ ,  $d$ , and  $e$  are nonzero, but they can be positive or negative.
- (b) Find two independent conservation laws for these reactions.

### Section 3.4

**3.11** For the given stoichiometric matrix, find the independent conservation laws, if any, determined using the Null Space Theorem.

$$(a) \quad \mathbf{S} = \begin{pmatrix} -1 & 3 \\ 1 & 2 \\ 0 & -1 \end{pmatrix} \quad (b) \quad \mathbf{S} = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 0 & 0 \\ 0 & -1 & -2 \end{pmatrix}$$

**3.12** This problem considers how to reconstruct the original reactions using the stoichiometric matrix and rate vector.

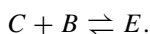
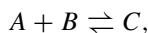
- (a) It is not possible to determine the original reactions using only the stoichiometric matrix. Demonstrate this by finding two different sets of reactions for

$$\mathbf{S} = \begin{pmatrix} -2 & 3 \\ 4 & -1 \\ 0 & 1 \end{pmatrix}.$$

- (b) Explain why it is not possible to determine the original reactions using only the rate vector.
- (c) Suppose that the rate vector is  $\mathbf{r} = (r_1, r_2)^T$ , where  $r_1 = k_1 AB^2$  and  $r_2 = CB$ . Using the stoichiometric matrix from part (a), determine the original reactions.

**3.13** This exercise considers some of the consequences of when all the reactions are reversible.

- (a) Suppose  $\mathbf{S}$  is  $m \times n$  and that  $\mathbf{S}$  has  $M$  independent rows. Explain why there are at least  $n - M$  independent conservation laws.
- (b) Consider the  $2N$  reactions, where  $N = 2$ ,



What is  $\mathbf{S}$ ? By simply looking at  $\mathbf{S}$ , and using the result from part (a), explain why there are at least  $N$  independent conservation laws. Also, explain why this is an immediate consequence of the reversibility of the reactions. Finally, show that the  $N$  conservation laws, when written as in (3.19), have positive coefficients.

**3.14** The stoichiometric matrix can be split into a reactant matrix  $\mathbf{A}$ , that contains the  $\alpha_{ij}$ 's in (3.50), and a product matrix  $\mathbf{B}$ , that contains the  $\beta_{ij}$ 's in (3.50). Doing this then  $\mathbf{S} = \mathbf{B} - \mathbf{A}$ . This problem explores how to use these matrices to find some of the steady states. You can assume that the matrices are  $4 \times 3$ , so there are four species and three reactions.

- (a) Suppose that the  $i$ 'th row of  $\mathbf{S}$  contains all zeros except for  $S_{i2}$ . Also, suppose that the  $i$ 'th row of  $\mathbf{A}$  is all zeros except for  $A_{i3}$ . Explain why, if there is a steady state, then it will have  $X_3 = 0$ .
- (b) Suppose that the  $i$ 'th row of  $\mathbf{S}$  has no negative entries. Assuming that  $S_{i2}$  is positive, suppose that the  $i$ 'th row of  $\mathbf{A}$  is all zeros except for  $A_{i3}$ . Explain why, if there is a steady state, then it will have  $X_3 = 0$ .

**3.15** Suppose  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k$  are the vectors formed from the coefficients for conservation laws for a set of reactions. Also, assume that  $k \geq 2$ .

- (a) If each law contains a species that does not appear in the other laws, what conclusion can you make about the entries of the  $\mathbf{a}_i$ 's?
- (b) Using your conclusion from part (a), show that the  $\mathbf{a}_i$ 's are independent.

- (c) Is the converse of the Independent Test true? Specifically, if the  $\mathbf{a}_i$ 's are independent, is it always possible to find an equivalent set of  $k$  vectors, formed from the coefficients for conservation laws for a set of reactions, so that each law contains a species that does not appear in the other laws?

### Section 3.5

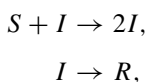
**3.16** For the systems below, find the steady states and determine if they are asymptotically stable. If present,  $b$  is a positive constant.

- (a)  $s' = c - s^2$   
 $c' = 1 + sc$   
 (b)  $u' = v - u$   
 $v' = (2 - u - v)(1 + v^2)$   
 (c)  $u' = v$   
 $v' = -b(1 - u^2)v + u$ ,  
 (d)  $x' = y - x$   
 $y' = 4x - y - xz$   
 $z' = xy - bz$ ,

**3.17** For the reactions below do the following: (1) write down the kinetic equations, (2) find independent conservation laws, (3) using the kinetic equations and the conservation law(s) determine the steady state, and (4) use the Linear Stability Theorem to determine the stability of the steady state. If the theorem does not cover the problem, state that the test is inconclusive. Also, assume that all reactants start out with a nonzero concentration.

- (a)  $2A \rightarrow 3Y + Z$   
 $X \rightarrow 2Y + A$   
 (c)  $S + S \rightarrow 2B$   
 $C \rightarrow S$   
 $S \rightarrow B$   
 (e)  $Y \rightarrow 2A + Z$   
 $3X \rightarrow A$   
 (b)  $S + B \rightarrow 2B$   
 $B \rightarrow A$   
 $S \rightarrow B$   
 (d)  $X \rightarrow 2Y + Z$   
 $A \rightarrow X + 3Z$   
 $Y \rightarrow 4Z$   
 (f)  $A + 2B \rightarrow 3B$   
 $A \rightarrow B$   
 $X \rightarrow A$   
 $B \rightarrow C$

**3.18** One method to reduce the spread of a disease is to vaccinate those who are susceptible. A model that attempts to account for the vaccination of the susceptible group is





Here  $S$ ,  $I$ , and  $R$  are the same groups used in the SIR model described in Sect. 3.3.2. Assume in the problem that  $S(0) = S_0$ ,  $I(0) = I_0$ ,  $R(0) = 0$ , where  $S_0$  and  $I_0$  are positive.

- Which reaction accounts for vaccinations? Explain why.
- Using the Law of Mass Action, write down the initial value problem that comes from the above reactions. After this, use a conservation law to reduce the kinetic equations down to a problem only involving  $S$  and  $I$ .
- Explain why the solution must satisfy  $0 \leq S \leq N$  and  $0 \leq I \leq N$ , where  $N = S_0 + I_0$ .
- What are the steady states for the reduced problem? What restrictions, if any, do you need to impose so the steady states satisfy the conditions in part (c)?
- One of the steady states you found has  $I = 0$ . Under what conditions is this steady state asymptotically stable?
- One of the steady states you found has  $I \neq 0$ , which is called an epidemic equilibrium. Under what conditions is this steady state asymptotically stable?
- With the long-term objective of keeping the number of infected individuals down to a minimum, what conditions, if any, should be imposed on the vaccination rate constant?

**3.19** This problem considers what is known as the standard SIR model with vital dynamics. Using the same three groups as in the SIR model, the kinetic equations are

$$\begin{aligned}\frac{dS}{dt} &= m(I + R) - \beta IS, \\ \frac{dI}{dt} &= \beta IS - (m + g)I, \\ \frac{dR}{dt} &= gI - mR.\end{aligned}$$

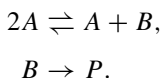
In this model it's assumed that members of the groups  $I$  and  $R$  die and are replaced with members in  $S$ . In what follows, assume that  $S(0) = S_0$ ,  $I(0) = I_0$ ,  $R(0) = 0$ , where  $S_0$  and  $I_0$  are positive.

- Show that the above system of equations can be derived from the Law of Mass Action, but this requires an assumption on the birth and death rates.
- Use a conservation law to reduce the above system to a problem for just  $S$  and  $I$ .
- Explain why the solution must satisfy  $0 \leq S \leq N$  and  $0 \leq I \leq N$ , where  $N = S_0 + I_0$ .

- (d) What are the steady states for the reduced problem in part (b)? What restrictions, if any, do you need to impose so the steady states satisfy the conditions in part (c)?
- (e) One of the steady states you found has  $I = 0$ . Under what conditions on the parameters is this steady state asymptotically stable?
- (f) One of the steady states you found has  $I \neq 0$ , what is called an epidemic equilibrium. Under what conditions on the parameters is this steady state asymptotically stable?
- (g) For measles  $m = 0.02$ ,  $\beta = 1800$ , and  $g = 100$  (Engbert and Drepper 1994). Show that in this case the epidemic equilibrium is asymptotically stable.

### Sections 3.6 and 3.7

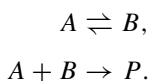
**3.20** A version of what is known as the Lindeman model for unimolecular reactions involves the following three reactions:



Assume that  $A(0) = A_0$  is nonzero,  $B(0) = 0$ , and  $P(0) = 0$ .

- (a) Using the Law of Mass Action, write down the initial value problem that comes from the above reactions.
- (b) Using a conservation law, explain why the kinetic equations can be reduced to just solving for  $A$  and  $B$ . Also, there is one steady state. What is it?
- (c) What conclusion, if any, can you make from the Linearized Stability Theorem?
- (d) Nondimensionalize the kinetic equations for  $A$  and  $B$  by using  $A_0$  to scale both  $A$  and  $B$ , and use  $t_c = 1/(k_1 A_0)$ . Use  $a$  and  $b$  as the nondimensional dependent variables. The final problem, including the initial conditions, should only contain the nondimensional parameters  $\lambda = k_2/(k_1 A_0)$  and  $\varepsilon = k_{-1}/k_1$ . It is going to be assumed that the reverse reaction  $2A \leftarrow A + B$  is slow compared to the forward reaction  $2A \rightarrow A + B$ , and this means we are assuming that  $\varepsilon \ll 1$ .
- (e) Find the first term in the expansions for  $a$  and  $b$ . Note that there is no layer in this problem.
- (f) Is the steady state found in part (b) unstable or asymptotically stable?

**3.21** Consider the reactions



Assume that  $A(0) = A_0$  is nonzero,  $B(0) = 0$ , and  $P(0) = 0$ .



- (a) Using the Law of Mass Action, write down the initial value problem that comes from the above reactions. After this use a conservation law to reduce this to a problem for just  $A$  and  $B$ .
- (b) There is one steady state. What is it?
- (c) Nondimensionalize the reduced problem in part (a), taking  $A(t) = A_0 a(\tau)$ ,  $B(t) = A_0 b(\tau)$ , and  $t = t_c \tau$ , where  $t_c = 1/(k_2 A_0)$ . The final problem, including the initial conditions, should contain the nondimensional parameters  $\lambda = k_{-1}/k_1$  and  $\varepsilon = k_2 A_0/k_1$ . Note that  $t_c$  is the time scale associated with the last reaction. It is going to be assumed that this reaction is slow compared to the other two, and this means we are assuming that  $\varepsilon \ll 1$ .
- (d) Find the first term in the outer expansions for  $a$  and  $b$ . Hint: You will need to consider the problem for the second term in the expansions to complete the determination of the first terms.
- (e) For the initial layer, find the first term in the expansions for  $a$  and  $b$ , and then match them with the outer solutions.
- (f) Determine a composite approximation for both  $a$  and  $b$ .
- (g) Is the steady state found in part (b) unstable or asymptotically stable?

**3.22** The Schnakenberg model for a chemical oscillator consists of the following two rate equations:

$$\begin{aligned} U' &= \mu^* - k_1 U V^2, \\ V' &= -k_2 V + k_1 U V^2, \end{aligned}$$

where  $\mu^*$  is a positive constant.

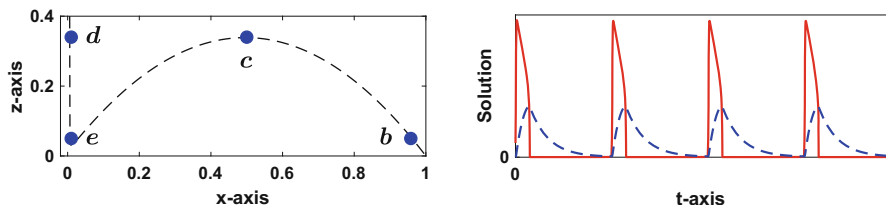
- (a) The  $k_i$  terms come from the Law of Mass Action. Find the two reactions that give rise to these three terms. For the record, the  $\mu^*$  term accounts for a constant influx of  $U$  into the system.
- (b) Show that the equations can be nondimensionalized to have the form

$$\begin{aligned} u' &= \mu - uv^2, \\ v' &= -v + uv^2, \end{aligned}$$

where  $\mu$  is a positive constant.

- (c) Using the equations from part (b), find the steady state and show that it is asymptotically stable if  $\mu > 1$  and it is unstable if  $\mu < 1$ .
- (d) Explain why the change in the stability as  $\mu$  decreases, and passes through  $\mu = 1$ , has the properties of a Hopf bifurcation as described in Example 3 of Sect. 3.5.3.

**3.23** The  $z$ -nullcline for (3.114) and (3.115) is shown in Fig. 3.23, on the left, and the solution curves are shown on the right.



**Fig. 3.23** Graphs for Exercise 3.23. Note that the points  $e$  and  $c$  are the local minimum and maximum, respectively, on the nullcline

- Assuming small  $\alpha$ , find first term approximations for the coordinates of the points  $e$  and  $c$ . Assume that  $f$  is independent of  $\alpha$ .
- As in part (a), find first term approximations for the coordinates of the points  $b$  and  $d$ . Note that the  $z$ -coordinate for  $d$  and  $c$  are the same, and the  $z$ -coordinate for  $b$  and  $e$  are the same.
- Explain why the closer  $\varepsilon$  gets to zero, the more the points  $b$ ,  $c$ ,  $d$ , and  $e$  determine the limit cycle, assuming there is a limit cycle solution.
- The solution curves in Fig. 3.23 are for a small  $\varepsilon$  and  $\alpha$ . Identify which is  $x(t)$  and which is  $z(t)$ . Also, locate the points  $b$ ,  $c$ ,  $d$ , and  $e$  in this graph. With this derive approximations for the amplitudes of the two functions.
- Explain why the  $x$ -coordinate of points  $c$  and  $e$  does not depend on  $f$ . Find a two-term expansion, for small  $\alpha$ , of the  $x$  coordinate for each of these points. For each point, use (3.120) to find a two-term expansion for  $f$  that results in the two nullclines intersecting. Use this to find a two-term expansion for the interval for  $f$  that produces a limit cycle solution.

## Sections 3.8

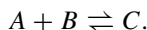
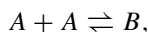
**3.24** A version of the Rozenzweig-MacArthur predator-prey model is

$$\begin{aligned}\frac{dS}{dt} &= \lambda S - \frac{\mu SP}{1 + \alpha S}, \\ \frac{dP}{dt} &= \frac{\beta SP}{1 + \alpha S} - \gamma P,\end{aligned}$$

where  $\lambda$ ,  $\mu$ ,  $\alpha$ ,  $\beta$ , and  $\gamma$  are positive constants. Because of the  $1 + \alpha S$  term these equations do not appear to be the direct application of the Law of Mass Action. However, this term is similar to what is obtained for Michaelis-Menten when using the QSSA. Derive a reaction scheme that produces the above equations when one of the species is assumed to be at a quasi-steady state.

**3.25** Although trimolecular reactions are rare in the real world, it is not uncommon to find trimerizations. These are reactions in which a product is constructed using three reactant molecules of the same species, with an effective overall reaction  $3A \rightarrow \text{products}$ . This exercise explores how to obtain this result using elementary reactions.

(a) One possible mechanism is



What are the resulting kinetics equations?

(b) Using the QSSA, and extreme parameter values, show how to reduce the equations in part (a) to obtain the approximate equation  $C' = -kA^3$ .

**3.26** In applying the QSSA to Michaelis-Menten one finds that the product's concentration satisfies an equation of the form

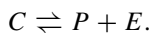
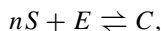
$$\frac{dP}{dt} = \frac{aS}{b+S},$$

where  $a$  and  $b$  are constants. It has been observed that in some reactions the product appears to follow a rate equation more of the form

$$\frac{dP}{dt} = \frac{aS^n}{b+S^n}.$$

In biochemistry this is known as Hill's equation and  $n$  is the Hill coefficient. This exercise explores how to obtain this result using the Law of Mass Action.

(a) One explanation is that  $n$  substrate molecules must get together with the enzyme to construct an intermediate complex  $C$ . This is the idea underlying cooperativity, and the reactions are

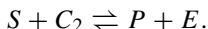
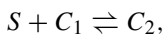
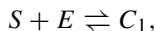


What are the resulting kinetic equations?

(b) Using the QSSA, and extreme parameter values, show how to reduce the equations in part (a) to obtain Hill's equation.

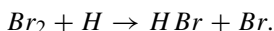
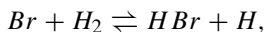
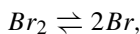
(c) The reactions in (a) are commonly assumed, but they require the unrealistic assumption that  $n+1$  molecules collide to form  $C$ . A more plausible explanation

is they interact sequentially. For  $n = 3$  the reactions are



Using the QSSA, and extreme parameter values, show how to reduce the kinetic equations to obtain Hill's equation.

**3.27** It is found experimentally that in the hydrogen-bromine reaction, the rate for the overall reaction of producing HBr from  $H_2$  and  $Br_2$  is  $r = kH_2Br_2^{1/2}$ . The implication is that the reaction is  $H_2 + \frac{1}{2}Br_2 \rightarrow HBr$ . This exercise explores one of the proposals for how this reaction proceeds as a sequence of elementary reactions. The assumption is

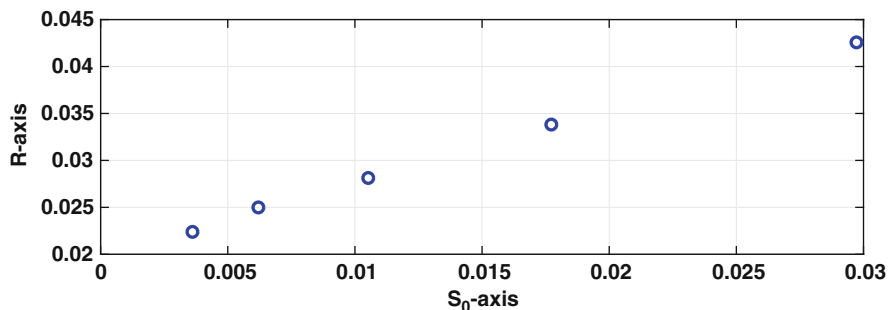


- What are the resulting kinetics equations?
- Using the QSSA, and extreme parameter values, show how to reduce the equations in part (a) so the resulting rate is  $r = kH_2Br_2^{1/2}$ .
- It is found, using SI units, that  $k_1 = 3.8 \times 10^{-8}$ ,  $k_{-1} = 4.2 \times 10^{-13}$ ,  $k_2 = 380$ ,  $k_{-2} = 7.2 \times 10^9$ , and  $k_3 = 9.6 \times 10^{10}$ . Are your assumptions in part (b) consistent with these values?

### *Additional Questions*

**3.28** This problem considers the proof of the Special Cases Stability Theorem given on page 130.

- Prove statement (i) directly from the Linearized Stability Theorem.
- Find out what the Hurwitz matrix is, and then write it down for (3.74) in the case of when  $n = 3$ . After this, find the determinants of the three leading principal minors of this matrix. According to the Hurwitz criteria, the roots of (3.74) have negative real part if the three determinants are positive. Moreover, it will have at least one root with positive real part if any of the three determinants are negative. Use this to prove statement (ii) of the theorem.
- Write down Descartes' rule of signs for positive roots, as it applies to (3.74). Use this to prove statement (iii) of the theorem.



**Fig. 3.24** Data for the hydrolysis of urea by the enzyme urease (Kryatov et al. 2000), where  $R = -S_0/S'(0)$ . In this graph, concentrations are measured in moles and time is in seconds

**3.29** Enzymatic reactions are characterized using  $v_M$  and  $K_M$ , as given in (3.88). For example, values of these constants are standard entries in biochemistry tables, such as Schomburg and Stephan (1997). This problem examines how they are used in conjunction with experimental data.

(a) Assuming the QSSA is valid, show that

$$\frac{1}{v_M} (S_0 + K_M) = -\frac{S_0}{S'(0)}.$$

The graph of the left-hand side of the above equation, as a function of  $S_0$ , produces what is known as a Hanes-Woolf plot. Given the experimental measurement of  $S_0/S'(0)$ , then linear regression can be used to determine  $1/v_M$  and  $K_M$ .

(b) Use the data in Fig. 3.24 to estimate  $v_M$  and  $K_M$ .

**3.30** This problem considers how to determine experimentally the stoichiometric coefficients and rate constant for a reaction. The specific reaction is given in (3.12). Also, different experiments correspond to using different initial concentrations.

- Suppose the initial velocity  $A'(0)$  can be determined experimentally. Explain how, using the results from three different experiments, to find  $\alpha$ ,  $\beta$ , and  $k$ .
- Suppose  $A'(0)$  and  $B'(0)$  are measured. Is it possible to determine  $\alpha$ ,  $\beta$ , and  $k$  using fewer than three experiments?
- Suppose it is possible to measure the concentrations as the reaction proceeds. Explain how to use the measurements at a specific time, say  $t = t_d$ , to determine  $\gamma$  and  $\delta$ . Assume that  $\alpha$ ,  $\beta$ , and  $k$  are known. Also, you should not have to solve the kinetic equations to find  $\gamma$  and  $\delta$ .

# Chapter 4

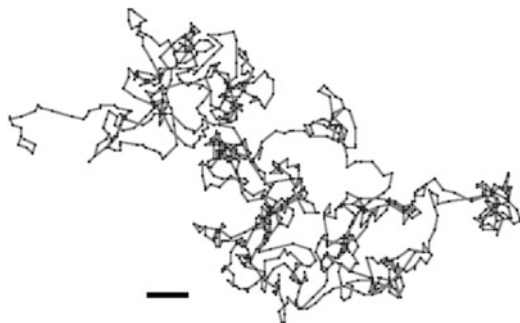
## Diffusion



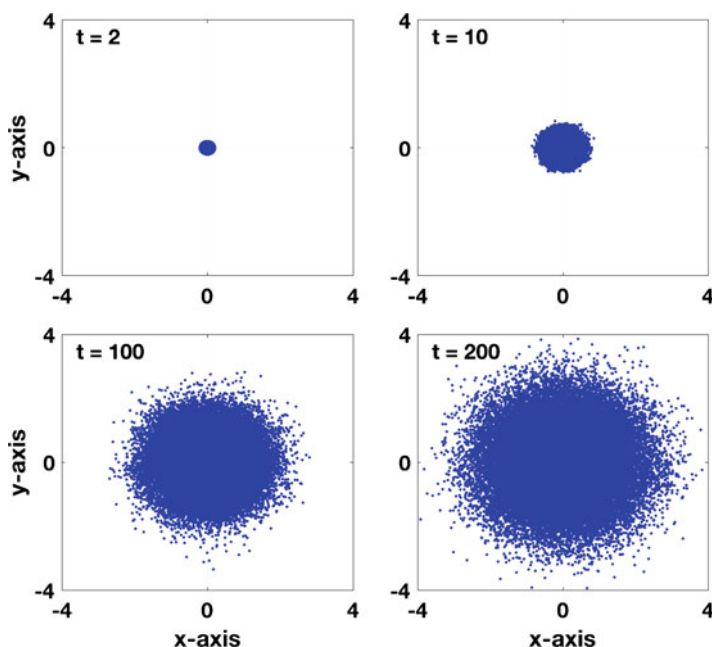
### 4.1 Introduction

In the last chapter we examined how to use the kinetics of reactions to model the rate of change of populations, or concentrations. We did not consider the consequences of the motion or spatial transport of these populations. There are multiple mechanisms involved with transport, and in this chapter we will examine one of them, and it is the process of diffusion. A simple example of diffusion arises when a perfume bottle is opened. Assuming the air is still, the perfume molecules move through the air because of molecular diffusion.

One mechanism responsible for diffusion is Brownian motion. Although the random microscopic movements associated with Brownian motion were observed as early as 1785, the first significant scientific study began with Robert Brown. In the summer of 1827 he made microscopic observations of granules from pollen that are suspended in water. What he saw surprised him as the tiny granules were in constant motion, never appearing to slow or stop, and following irregular paths much like the one in Fig. 4.1. Moreover, he found that this motion was not caused by external influences such as light or convection currents. He also quickly ruled out the possibility that the granules were somehow alive. However, the underlying reasons for the movement remained elusive. It was not until the early 1900s that the theoretical work of Einstein, and the experimental work of Perrin, explained the motion. What is happening is that the granules, which are approximately  $2\text{ }\mu\text{m}$  in length, are under constant bombardment by the surrounding water molecules. Although the latter are much smaller, having a diameter of approximately  $3 \times 10^{-4}\text{ }\mu\text{m}$ , there are many of them and they are responsible for a very large number of random impacts on each granule. The irregular nature of this forcing gives rise to the randomness of the motion. It is now known that Brownian fluctuations are essential to widely diverse phenomena, from passive transport of ions and nutrients for biological cells to models for financial assets.



**Fig. 4.1** The path recorded for a micron-sized silica particle due to Brownian motion over a 2.2 s interval. The black bar is 10  $\mu\text{m}$  (Blum et al. 2006)



**Fig. 4.2** Scatter plots showing the distribution of 50,000 randomly moving particles, which start out at the origin. The distributions are shown at  $t = 2, 10, 100$ , and  $200$

If one starts with a large number of nearby particles, each undergoing Brownian motion, then over time the particles will tend to be spread throughout the medium. This is illustrated in Fig. 4.2, which shows the positions of 50,000 randomly moving particles that all start out at the origin (the formulas used to generate these plots are given in Sect. 4.7). The change in the particle density is an example of diffusion. Thus, diffusion can be thought of as a macroscopic manifestation of the random motion that is taking place on the microscopic level. This observation also

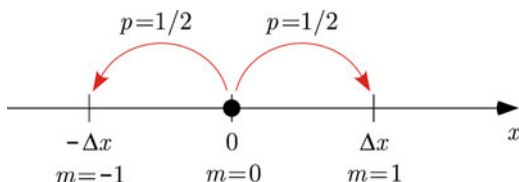
identifies our approach in modeling diffusion. We will start with random walks at the microscopic scale and show how they can give rise to diffusion on the macroscopic scale. This differs from the more classic approach, considered later in the chapter, of deriving the diffusion equation using a balance law. The random walk approach provides an explanation of the possible underlying mechanisms in diffusion, and so we will begin with it.

## 4.2 Random Walks and Brownian Motion

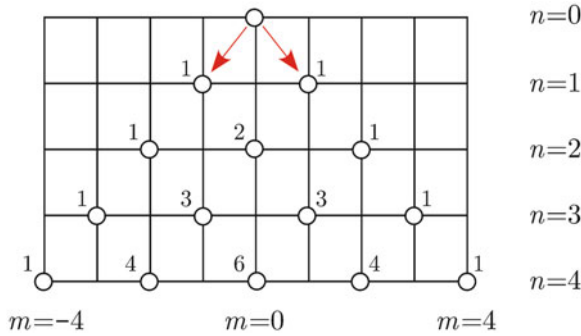
The rapid fluctuating movement of a molecule is the result of impacts with the surrounding atoms and molecules. To construct a mathematical model for this situation we will consider the motion to be one-dimensional. Specifically, the molecules move back and forth along the  $x$ -axis. To account for the randomness of the motion, consider a single molecule that starts out at  $x = 0$ . After a time step  $\Delta t$ , the molecule moves a distance  $\Delta x$  either to the right or left and it moves in either direction with equal probability. One way to think of this is that the molecule has a coin, flips it, and based on whether the outcome is heads or tails it moves left or right. A diagram illustrating the choices, and outcomes, is shown in Fig. 4.3. At time step  $n = 1$  the coin is again flipped and this determines whether the molecule will step left or right at time level  $n = 2$ . The various positions that are possible when starting at  $(x, t) = (0, 0)$  are shown in Fig. 4.4. Also included in this figure is the number of paths that are available to reach the respective point. For example, to reach  $m = -1$  at  $n = 3$  there are three possible paths. Letting L designate a left move, and R a right move, the three paths are LLR, LRL, and RLL. It is also evident that this number is the sum of the paths for the two adjacent positions,  $m = -2$  and  $m = 0$ , at the previous time step.

Carrying out the above procedure the molecule moves back and forth on the  $x$ -axis and the path it follows is called a random walk. One observation concerning Fig. 4.3 is that at any given time level not all spatial positions are possible. For example, it is impossible for the molecule to be at  $x = 0$  when  $n = 1, 3, 5, \dots$ . Also, each step in the random walk is independent of the preceding one. This lack of memory is characteristic of what is known as the Markov property. Finally, note that the number of paths shows more than a passing resemblance to Pascal's triangle. This connection will not be used in what follows because we will be interested in generalizations of this problem that do not have a Pascal's triangle structure.

**Fig. 4.3** The first step in a random walk. In going from time step  $n = 0$  to  $n = 1$ , the molecule moves a distance of  $\Delta x$  to the right or left with equal probability







**Fig. 4.4** The positions that are possible in the random walk are indicated by the circles. The molecule starts at  $(m, n) = (0, 0)$ , and the two red arrows indicate where the particle can move (the arrows for the other positions are not shown). The number next to each circle is the number of unique paths that are available to reach that location, when starting at  $(0, 0)$

We want to keep track of the molecule's position and given the way it is determined it should not be unexpected that probabilistic methods are needed. With this in mind, let  $w(m, n)$  be the probability that the molecule is at  $x = m\Delta x$  after  $n$  time steps. The time steps have a fixed value  $\Delta t$ , so, after  $n$  steps  $t = n\Delta t$ . In preparation to calculating  $w$  it is worth stating a few of the more interesting properties of this function that are evident from Fig. 4.4.

- Given any time level  $n$ , the points where  $w$  is nonzero are  $m = -n, -n + 2, \dots, n - 2, n$ . It is zero at all other values of  $m$ .
- The number of paths available to reach  $x = m\Delta x$ , at time step  $n$ , is equal to the number of available paths to reach  $m - 1$ , at time step  $n - 1$ , added to the number to reach  $m + 1$ , at time step  $n - 1$ .

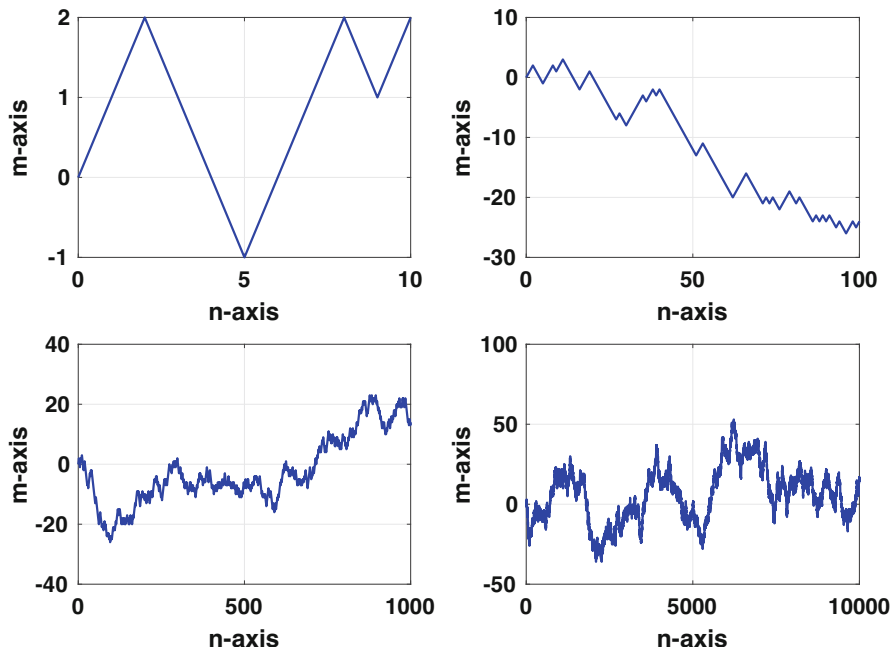
This conclusion is a consequence of the observation that to be able to be located at  $x = m\Delta x$  at  $t = n\Delta x$ , it is necessary to be located to either  $x = (m + 1)\Delta x$  or  $x = (m - 1)\Delta x$  at time  $t = (n - 1)\Delta x$ .

- Because all paths are equally likely,

$$w(m, n) = \frac{\text{number of paths from } (x, t) = (0, 0) \text{ to } (x, t) = (m\Delta x, n\Delta t)}{\text{total number of paths from } t = 0 \text{ to } t = n\Delta t} \quad (4.1)$$

- The total number of paths from  $t = 0$  to  $t = n\Delta t$  is  $2^n$ .  
The reason this holds is that a particle has two potential paths to the next time level. Hence, the total number of paths doubles with each time step.
- At each time step  $n$  the molecule must, with probability one, be located somewhere along the  $x$ -axis. In other words,  $\sum_{m=-\infty}^{\infty} w(m, n) = 1$ .

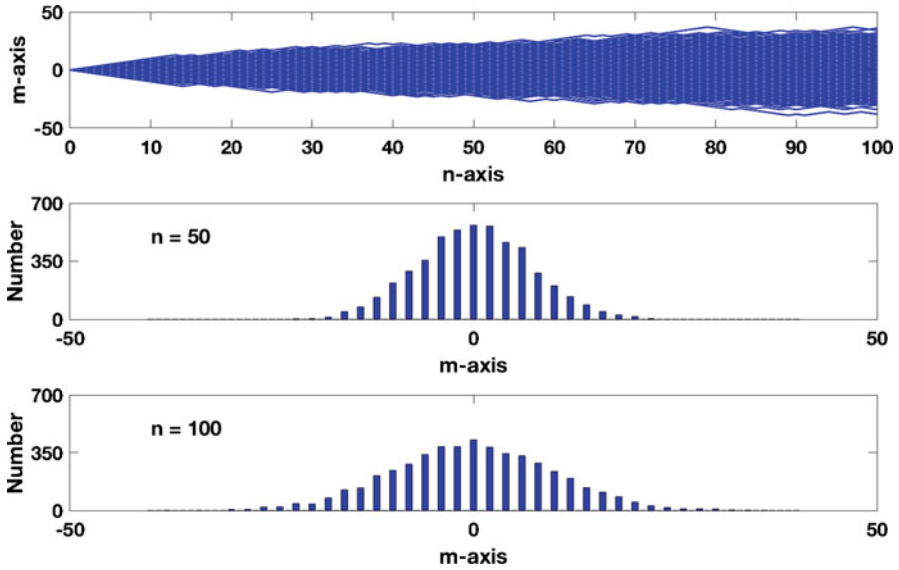
It is not difficult to write a computer program for a random walk, and to then use this to investigate what happens when one uses a rather large number of time steps. An example is shown in Fig. 4.5. The zigzag nature of a random walk is clearly



**Fig. 4.5** Example of a random walk shown over successively longer time intervals

seen in all four graphs, and as  $n$  increases the path shows a large-scale variation in conjunction with the small-scale jagged motion due to the random changes in direction. Another point to make is that if the calculation is redone, a very different path would appear. The reason is that when  $n = 10,000$  there are  $2^{10,000} \approx 10^{3010}$  different possible paths. Consequently, it is statistically zero that the same path would be obtained.

Another experiment of interest is when a large number of random walks are run, each starting at  $(0, 0)$ . The outcome of one such experiment, using 5000 walks, is shown in Fig. 4.6. Looking at the cross-sectional profiles, at  $n = 50$  and  $n = 100$ , it is seen that the profiles resemble bell curves (also known as a normal or Gaussian distributions). The symmetry in  $m$  is due to the equal probability of jumping right or left. Also, even though the positions are spreading out as  $n$  increases, no particle has come close to reaching the maximum possible distance from  $m = 0$ . This is not really expected using only 5000 paths because the probability of being at  $m = \pm 100$ , when  $n = 100$ , is  $2^{-99} \approx 10^{-67}$ . The objective of the next section is to derive an analytical approximation of the cross-sectional profiles seen in Fig. 4.6.



**Fig. 4.6** Top: 5000 random walk paths, all starting at  $m = 0$ . The lower two graphs show the spatial distribution of the particles at  $n = 50$ , and then at  $n = 100$

### 4.2.1 Calculating $w(m, n)$

As stated in (4.1), because all paths are equally likely, the probability of being located at  $x = m \Delta x$  at time step  $n$  is equal to the number of unique paths available to reach this point divided by the total number of paths available to reach time level  $n$ . For example, in Fig. 4.4, there are six paths that are able to reach  $x = 0$  at time step  $n = 4$ . Because the total number of paths is  $2^4$  it follows that  $w(0, 4) = 6/2^4 = \frac{3}{8}$ . We will use this observation to determine the general formula for  $w(m, n)$ .

To reach any point after  $n$  time steps requires a sequence of left and right spatial steps. If  $q$  is the number of left steps, then the number of right steps is  $n - q$ . One way to think of this is that you have  $q$  of the  $L$ 's, along with  $n - q$  of the  $R$ 's, and these are going to be arranged in an  $n$ -vector. Picking one of the  $L$ 's, there are  $n$  positions it can be placed in the vector, while for the second  $L$  there are  $n - 1$  positions, etc. The result is that the total number of choices we can make is  $n(n - 1)(n - 2) \cdots (n - q + 1)$ . For example, in regard to Fig. 4.4, reaching  $m = 0$  at  $n = 4$  requires two  $L$ 's and two  $R$ 's. The unique orderings we can have are  $LLRR$ ,  $LRLR$ ,  $LRRL$ ,  $RLLR$ ,  $RLLR$ , and  $RRLL$ . However, because  $n = 4$  and  $q = 2$  our formula states that the number should be  $4 \times 3 = 12$ . The reason for the discrepancy is because we have considered the  $L$ 's as distinct from each other. Thinking this way we would conclude that two  $L$ 's, say  $L_1$  and  $L_2$ , could produce two different paths,  $L_1 L_2$  and  $L_2 L_1$ . They do not and therefore we must divide by  $2!$ , or in the general case by  $q!$ . Because there are  $2^n$  paths in total it therefore follows that

$$\begin{aligned}
 w(m, n) &= \frac{n(n-1)(n-2) \cdots (n-q+1)}{2^n q!} \\
 &= \frac{n!}{2^n q!(n-q)!} \quad \text{for } m = -n, -n+2, \dots, n-2, n.
 \end{aligned} \tag{4.2}$$

To relate the value of  $q$  with the spatial position note that if there are  $q$  moves to the left, and  $n - q$  moves to the right, then  $x = -q\Delta x + (n - q)\Delta x$ . Because we also have that  $x = m\Delta x$  it follows that  $m = n - 2q$ , or equivalently,

$$q = \frac{1}{2}(n - m). \tag{4.3}$$

We have done what we set out to do, which is to produce a formula for  $w(m, n)$ . It is not obvious what a plot of  $w$  would look like although we can anticipate some of the major features. Because it is equally likely to move left or right the plot of  $w$ , as a function of  $m$ , should be symmetric about  $m = 0$ . For the same reason, it is expected that  $w$  decreases with distance from  $m = 0$ . To support these observations,  $w$  is plotted in Fig. 4.7 for two values of  $n$ . It shows the expected behavior but also notice the spreading of the peak as  $n$  increases and the corresponding drop in the maximum value at  $m = 0$ . This is very typical for a solution that is describing a process controlled by diffusion. It is also not a coincidence that the curves in Fig. 4.7 have the same structure as the distribution curve in Fig. 4.6. We will return to this observation when the point source solution is discussed later in the chapter.

Random walks are used in a wide variety of applications and because of this the terminology varies a bit with the area. For example, they are used in gas dynamics to describe the motion of atoms in a gas. Such atoms do not travel in a straight line, but rather undergo random changes of direction due to frequent collisions with other atoms. This is modeled as a random walk where the spatial jump  $\Delta x$  is called the mean free path and it is a measure of the distance that an atom travels between two successive collisions. As an estimate of this distance, at room temperature the mean

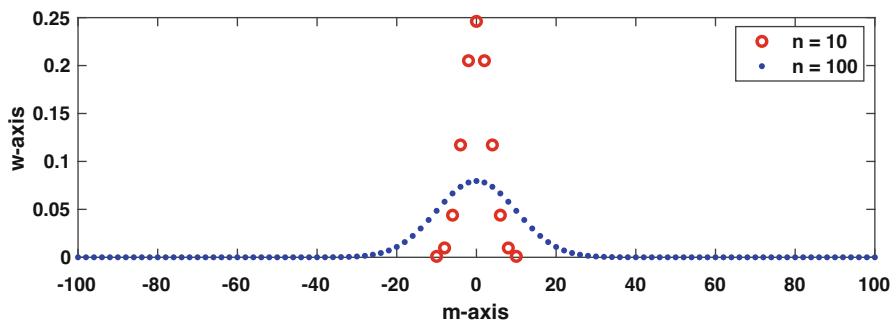


Fig. 4.7 The nonzero values of  $w(m, n)$ , as given in (4.2), as a function of  $m$

free path in air is about  $10^{-7}$  m and a typical molecule undergoes up to  $10^9$  collisions per second. This small spatial scale, and the enormous number of collisions, are the basis for the continuous limit considered later.

### 4.2.2 Large $n$ Approximation

Given the regular structure of the function in Fig. 4.7 it would be worthwhile to see if we can simplify our formula to make it a bit easier to work with. The assumption we need to pull this off is that  $n$  is large. If this is the case we can make use of Stirling's approximation for the factorial, which is

$$n! \sim e^{-n} n^n \sqrt{2\pi n} \left(1 + \frac{1}{12n} + O\left(\frac{1}{n^2}\right)\right). \quad (4.4)$$

It will be assumed that not only is  $n$  large, but the number of left moves and the number of right moves are also large. Using the first term in Stirling's approximation for each of the factorials in (4.2) we get

$$w(m, n) \sim \frac{n^n}{2^n q^q (n-q)^{n-q}} \sqrt{\frac{n}{2\pi q(n-q)}}.$$

Recalling that  $q = (n-m)/2$  and  $n-q = (n+m)/2$ , then the above approximation can be written as

$$w(m, n) \sim \sqrt{\frac{2n}{\pi(n+m)(n-m)}} Q, \quad (4.5)$$

where

$$Q = \left(\frac{n}{n+m}\right)^{(n+m)/2} \left(\frac{n}{n-m}\right)^{(n-m)/2}. \quad (4.6)$$

It is assumed here that  $m \neq \pm n$ .

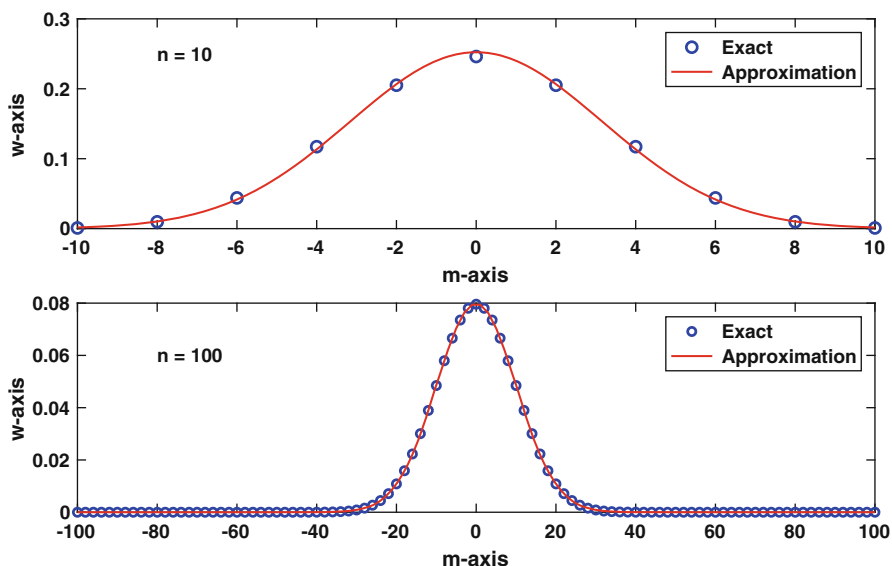
We can simplify  $Q$  using the assumption that  $n$  is large, or more specifically that  $m/n$  is small. Both factors in (4.6), for large  $n$ , fall into the category of an indeterminate form of type  $1^\infty$ . In calculus the method used to analyze such expressions involves taking the natural log of the function. Doing this, and using the Taylor expansion of  $\ln(1+x)$  given in Table 2.1, we obtain the following:

$$\begin{aligned}
\ln(Q) &= \frac{n+m}{2} \ln\left(\frac{n}{n+m}\right) + \frac{n-m}{2} \ln\left(\frac{n}{n-m}\right) \\
&= -\frac{n+m}{2} \ln\left(1 + \frac{m}{n}\right) - \frac{n-m}{2} \ln\left(1 - \frac{m}{n}\right) \\
&\sim -\frac{n+m}{2} \left[ \frac{m}{n} - \frac{1}{2} \left(\frac{m}{n}\right)^2 + \dots \right] - \frac{n-m}{2} \left[ -\frac{m}{n} - \frac{1}{2} \left(\frac{m}{n}\right)^2 + \dots \right] \\
&= -\frac{1}{2} \frac{m^2}{n} + \dots
\end{aligned} \tag{4.7}$$

With this, we conclude that, for small  $m/n$ , the nonzero values of  $w$  can be approximated as

$$w(m, n) \sim \sqrt{\frac{2}{\pi n}} e^{-m^2/(2n)}. \tag{4.8}$$

This expression is significantly simpler than (4.2). To examine the accuracy of this approximation, in Fig. 4.8 this function is plotted along with the exact values for two different values of  $n$ . It is seen that even in the case of  $n = 10$  the approximation is rather good, and it improves significantly when  $n$  is larger. It is also evident that (4.8) provides a reasonable approximation on the far left and right, regions where  $m/n$  is not particularly small.



**Fig. 4.8** Comparison between the exact nonzero values of  $w(m, n)$  calculated using (4.2), and the approximation in (4.8)

### 4.3 Continuous Limit

As formulated in the last section, a random walk involves discrete steps in space and time. We are now going to investigate the situation when the number of time steps becomes so large that the process is effectively a continuous function of time. As we do this it will be necessary to adjust the spatial stepsize  $\Delta x$ , but we will wait and let the analysis tell us just how to do this. To set the stage, we fix the time interval, and so, it is assumed  $0 \leq t \leq T$ . Of interest is what happens to the random walk solution as we use smaller and smaller time steps, thereby increasing the number of steps needed to reach  $t = T$ . It is assumed that the spatial steps are also getting smaller. One way to think about this is that the steps become so small that the motion takes on the appearance of a continuous function of time and space and not one that is making discrete jumps.

The starting point is Fig. 4.4. As pointed out earlier, for this grid the number of paths available to reach  $x = m\Delta x$  at time step  $n$  is equal to the number of the paths to reach  $m - 1$  added to the number to reach  $m + 1$  at time step  $n - 1$ . Writing this as

$$\text{paths for } (m, n) = \text{paths for } (m - 1, n - 1) + \text{paths for } (m + 1, n - 1),$$

then we have that

$$\begin{aligned} \frac{\text{paths for } (m, n)}{2^n} &= \frac{\text{paths for } (m - 1, n - 1)}{2^n} + \frac{\text{paths for } (m + 1, n - 1)}{2^n} \\ &= \frac{1}{2} \frac{\text{paths for } (m - 1, n - 1)}{2^{n-1}} + \frac{1}{2} \frac{\text{paths for } (m + 1, n - 1)}{2^{n-1}}. \end{aligned}$$

Using the function  $w(m, n)$ , the above equation takes the form

$$w(m, n) = \frac{1}{2} w(m - 1, n - 1) + \frac{1}{2} w(m + 1, n - 1). \quad (4.9)$$

This important result gives us a formula for the probability function, and it is the basis for what is called a *master equation* for a stochastic process.

To switch from  $m, n$  to  $x, t$  recall that  $x = m\Delta x$ . Also, if we are using  $n$  time steps to reach  $t = T$ , then the size of each step is  $\Delta t = T/n$ . By introducing the function  $u(x, t) = w(m, n)$ , then (4.9) can be written as

$$2u(x, t) = u(x - \Delta x, t - \Delta t) + u(x + \Delta x, t - \Delta t). \quad (4.10)$$

It remains to use Taylor's theorem for small  $\Delta x$ ,  $\Delta t$ . Expanding (4.10) up to the second-order yields the following:

$$\begin{aligned} 2u &= u - \Delta x u_x - \Delta t u_t + \frac{1}{2}(\Delta x^2 u_{xx} + 2\Delta x \Delta t u_{xt} + \Delta t^2 u_{tt}) + \dots \\ &\quad + u + \Delta x u_x - \Delta t u_t + \frac{1}{2}(\Delta x^2 u_{xx} - 2\Delta x \Delta t u_{xt} + \Delta t^2 u_{tt}) + \dots \\ &= 2u - 2\Delta t u_t + \Delta x^2 u_{xx} + \Delta t^2 u_{tt} + \dots \end{aligned}$$

In the above expression  $u$  and its derivatives are evaluated at  $(x, t)$ . Rearranging things a bit we obtain (see Exercise 4.9)

$$u_t = \frac{(\Delta x)^2}{2\Delta t} u_{xx} + \frac{\Delta t}{2} u_{tt} + \dots \quad (4.11)$$

The question is, what equation is obtained for small  $\Delta x$  and  $\Delta t$ ? As with the Goldilocks story, there are three possibilities and they are based on what happens to the ratio  $(\Delta x)^2/\Delta t$  as  $\Delta x$  and  $\Delta t$  approach zero. If  $(\Delta x)^2/\Delta t$  becomes unbounded, then the first term approximation we obtain from (4.11) is  $u_{xx} = 0$ . Given that  $u \rightarrow 0$  as  $x \rightarrow \pm\infty$  we conclude that  $u = 0$ . For the other extreme, when  $(\Delta x)^2/\Delta t \rightarrow 0$  as  $\Delta x$  and  $\Delta t$  approach zero, we obtain  $u_t = 0$ . This equation only applies to the steady state and it is unable to describe the time-dependent changes seen in the solution. The limit that is “just right,” what mathematicians call the distinguished limit, is the case of when  $(\Delta x)^2/\Delta t$  has a fixed value as  $\Delta x$  and  $\Delta t$  approach zero. For this reason, we will assume

$$D = \frac{(\Delta x)^2}{2\Delta t} \quad (4.12)$$

remains constant in the limit. In this case, we conclude from (4.11) and (4.12) that

$$u_t = D u_{xx}, \quad (4.13)$$

where the constant  $D$  is known as the *diffusion coefficient* for the problem. This is the diffusion equation. As derived,  $u(x, t)$  is a continuous approximation for the nonzero values of  $w(m, n)$ . One of the more interesting aspects of this is that it effectively provides a smooth macroscopic description of the random microscopic movements of the molecules.

### 4.3.1 What Does $D$ Signify?

The only parameter appearing in the diffusion equation is  $D$ , and its value signifies the strength or weakness of the underlying diffusion process. From its definition



in (4.12), it is seen that the larger the value of  $D$  the farther the molecules move in a given time step. In a medium where the molecules are more closely packed, so the random walk steps are not particularly large, the diffusion coefficient is not as big as it would be in a more dilute mixture. It is not surprising therefore that for a gas diffusing in air  $D \approx 10^{-5} \text{ m}^2/\text{s}$  while for a soluble material in water  $D \approx 10^{-9} \text{ m}^2/\text{s}$ .

Brownian motion, which arises when the rapid fluctuations in a particles position are caused by impacts from many surrounding smaller particles, was used to introduce the random walk model. However, random walks are applicable to other types of diffusion processes. What follows are some well-studied examples, and the corresponding interpretation of the diffusion coefficient in each case.

### Diffusion in a Gas

In a gas, molecules undergo a sequence of collisions with other molecules, giving rise to what is effectively a random walk. In this case, the spatial jump  $\Delta x$  is taken to be the average distance  $\lambda$  that the molecule travels before changing direction, what is known as the mean free path of the molecule. In conjunction with this,  $\Delta t$  is assumed to be equal to the average time  $\tau$  between collisions. With this, the diffusion coefficient (4.12) is written as

$$D = \frac{\lambda^2}{2\tau} . \quad (4.14)$$

This is often referred to as the Einstein-Smoluchowski equation. It is useful as it can be used to experimentally determine the value of  $D$ . For example, at room temperature,  $O_2$  is found to have a mean free path of 80 nm and an average speed  $v$  of approximately 400 m/s. Assuming  $v = \lambda/\tau$ , then  $D = 2 \times 10^{-5} \text{ m}^2/\text{s}$ . Perhaps a more interesting observation is that  $\tau = \lambda/v = 2 \times 10^{-10} \text{ s}$ , which means a molecule of  $O_2$  undergoes  $5 \times 10^9$  collisions per second. It should be pointed out that this is for one spatial dimension. As will be explained in Sect. 4.7, the three-dimensional version of (4.14) is  $D = \lambda^2/(6\tau)$ . Therefore, although the precise value of the diffusion coefficient is affected by dimension, the order of magnitude is not.

### Diffusion in a Fluid

Given that  $D$  is a measure of the ability of a molecule to move through the maze created by its neighboring atoms and molecules, it should not be surprising to learn that the larger the molecule the smaller the diffusion coefficient. The formula in (4.14), however, contains no information related to the structure or state of the molecule or its surrounding medium. It is possible to derive such information, and an example is the Stokes-Einstein equation

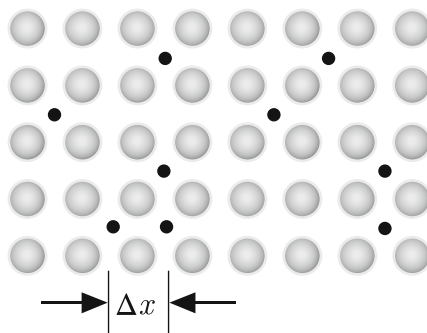
$$D = \frac{kT}{6\pi r\eta} . \quad (4.15)$$

This assumes the molecules are spheres, where  $T$  is the temperature in Kelvin,  $r$  is the radius of the molecule,  $\eta$  is the dynamic viscosity of the solvent, and  $k$  is known as the Boltzmann constant. How it is possible to get fluid viscosity into the formula for the diffusion coefficient will be explained in Sect. 4.8. The interest in (4.15), at this point, is the realization that the diffusion coefficient does depend on the size of the molecule, and decreases as the molecule's radius increases. It is also interesting what information can be derived from (4.15). For example, Einstein was able to use (4.15) to calculate Avogadro's number  $N_A$ , which is the number of molecules in a mole. From the kinetic theory for gases he knew that  $k = R/N_A$ , where  $R$  is the universal gas constant, and from this he rewrote (4.15) as  $N_A = RT/(6\pi r\eta D)$ . Using independent measurements of the constants on the right-hand side of this equation Einstein obtained a simple method for finding  $N_A$ .

### Diffusion in a Solid

Diffusion also occurs in solids, although the process is fundamentally different from what occurs in gases and liquids. One mechanism, known as interstitial diffusion, is illustrated in Fig. 4.9. Solids have a well-defined atomic structure and in metals the atoms generally form a lattice pattern. Smaller atoms are able to move through the solid by jumping between adjacent interstitial spaces. This requires the adjacent space to be unoccupied, and so this form of diffusion applies to dilute concentrations of diffusing atoms. Also, the diffusing atoms must be small enough to be able to make the jumps. For example, hydrogen, oxygen, nitrogen, and carbon are able to diffuse interstitially through metals, such as iron. However, the lattice points are relatively close so even small interstitial atoms must push their way through to the adjacent opening. This requires them to have sufficient energy to be able to squeeze through. It is possible to account for this in the diffusion coefficient by noticing that  $D = p\Delta x^2/\Delta t$ , where  $p$  is the probability of a jump (a more rigorous explanation of this can be found in Exercise 4.8). It is known that the probability of a successful jump depends on the thermal energy, and the higher the temperature, the greater the likelihood of a successful jump. Using reaction rate theory it has been found that the specific form is

**Fig. 4.9** Interstitial diffusion in a solid. The atoms of the solid form a lattice, and the smaller interstitial atoms move through the lattice by undergoing a random walk



$$D = D_0 e^{-E/(kT)}, \quad (4.16)$$

where  $E$  is the activation energy,  $k$  is the Boltzmann constant, and  $T$  is the absolute temperature. Also,  $D_0$  is the free solution diffusion coefficient, which is the value obtained when  $T \rightarrow \infty$ . This dependence on temperature is the basis for manufacturing hardened metals, where the metal is heated to allow diffusion of carbon through the lattice. For example, heating steel and allowing carbon to diffuse into the metal produces a much stronger surface, a process known as carburization. The complication is that this is very slow. The reason is that, even at  $900^\circ\text{C}$ , the diffusion coefficient for carbon is very small, on the order of  $10^{-11} \text{ m}^2/\text{s}$ .

Interstitial diffusion is involved in the operation of fuel cells, such as those in some hybrid vehicles, as well as is the production and operation of nanoelectronic devices. This is an active research area and those interested could consult Shaw (2017) and Tahini et al. (2015).

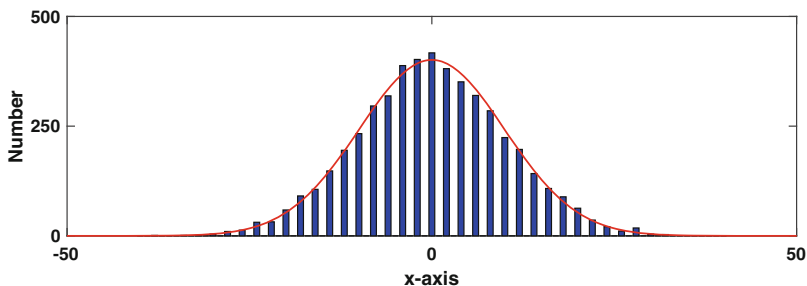
## 4.4 Solutions of the Diffusion Equation

Now that the diffusion equation has been derived, the next question to address is how to find the solution. This equation has been studied for almost two centuries, and this has given mathematicians time to find multiple ways to construct the solution. Some of the possibilities include separation of variables, and transform methods. The latter are used in the next section. What is considered now are the solutions we have already derived. As will illustrated in the examples to follow, although being able to find the solution is important, in applied mathematics it is equally important that you know how to use the solution. In many applications, the mathematical analysis is done in conjunction with experimental research, and it is necessary that the two reinforce each other. Demonstrating how this might be done is the objective of the examples below.

### 4.4.1 Point Source Solution

Given that the continuous approximation of the nonzero values for  $w(m, n)$  produces the diffusion equation, we will investigate what happens to the large  $n$  approximation of  $w(m, n)$  given in (4.8). Recalling that  $u(x, t) = w(m, n)$ ,  $x = m\Delta x$ , and  $t = n\Delta t$ , then

$$\begin{aligned} u(x, t) &\sim \sqrt{\frac{2\Delta t}{\pi t}} e^{-x^2 \Delta t / (2t \Delta x^2)} \\ &= \frac{\Delta x}{\sqrt{\pi D t}} e^{-x^2 / (4Dt)}. \end{aligned} \quad (4.17)$$



**Fig. 4.10** The outcome of 5000 random walks, all starting at  $x = 0$ , at time step  $n = 100$ . The solid curve is  $Pu(x, t)$ , where  $P = 5000$  and  $u(x, t)$  is given in (4.17) for  $\Delta x = \Delta t = 1$

It is not hard to verify that this function satisfies the diffusion equation in (4.13). To compare it with the random walk experiment, suppose  $P$  random walks are carried out, all starting at  $x = 0$ . An example of this is shown in Fig. 4.6, for  $P = 5000$ . The probability  $w(m, n)$  is determined experimentally by counting the number of paths that go through the point  $(m, n)$ , and then dividing this by  $P$ . Said another way,  $Pu(x, t)$  approximates the total number of paths that pass through  $(m, n)$ , assuming  $P$  is large and  $w(m, n)$  is nonzero. A confirmation of this is given in Fig. 4.10, which shows that  $Pu(x, t)$  does indeed provide an excellent approximation for the number of paths.

Each path in the above experiment represents the motion of a particle. In many applications, it is not the number of paths that are of interest but, rather, the resulting concentration of the particles. In this case, concentration means the number of particles per unit length. To determine this, recall that the nonzero values of  $w(m, n)$  are separated by a distance  $2\Delta x$ . Consequently, the concentration is  $Pw(m, n)/(2\Delta x)$ , or equivalently,  $Pu(x, t)/(2\Delta x)$ . Letting  $c(x, t)$  denote the concentration, then, from (4.17),

$$c(x, t) = P \frac{1}{2\sqrt{\pi Dt}} e^{-x^2/(4Dt)}. \quad (4.18)$$

This is known as the *point source solution* with strength  $P$ . More specifically, it is the solution when there is a point source of strength  $P$ , at  $x = 0$ , in the initial condition.

There are three conditions  $c(x, t)$  must satisfy to qualify as a point source solution. First, for  $t > 0$ , it must satisfy the diffusion equation

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2}. \quad (4.19)$$

It is straightforward to show that (4.18) satisfies this equation. Second, if the point source in the initial condition is located at  $x = 0$ , then  $c(x, t)$  must satisfy, for  $x \neq 0$ ,

$$\lim_{t \rightarrow 0^+} c(x, t) = 0. \quad (4.20)$$

It can be shown that (4.18) satisfies this condition using l'Hospital's rule. Third, if the source has strength  $P$ , then, for  $t > 0$ , the following must hold:

$$\int_{-\infty}^{\infty} c(x, t) dx = P. \quad (4.21)$$

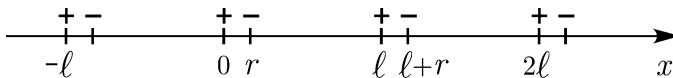
Physically this means that no matter the value of  $t$ , the total number of particles along the  $x$ -axis is  $P$ . Proving this is not hard, and uses the fact that, if  $a > 0$ , then

$$\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}}. \quad (4.22)$$

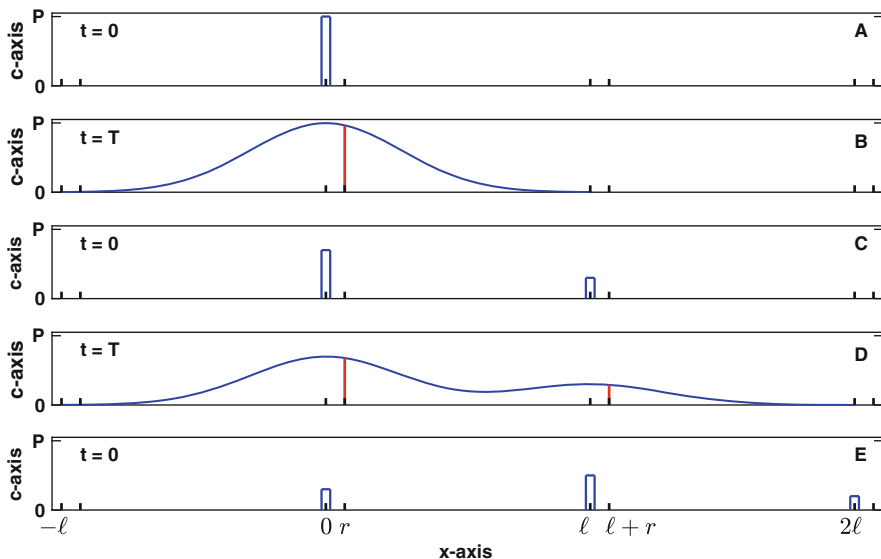
The solution in (4.18) is important enough that it has other names, depending on the context in which it arises. For example, when  $P = 1$ , it is also referred to as the heat kernel, or as the fundamental solution of the heat equation. It is also possible to write the associated initial condition using what is known as the Dirac delta function  $\delta(x)$ . Namely, for (4.18), it is  $c(x, 0) = P\delta(x)$ . Finally, it is possible to derive the point source solution directly by solving the diffusion problem using a similarity method. How this is done is outlined in Exercise 1.16.

*Example (Brownian Ratchet)* It is possible to take advantage of the Brownian motion of molecules. One application is based on what is known as a Brownian ratchet, where particles are moved in one direction by using a combination of diffusion and externally applied forces. This idea has been used to separate solutions of DNA, using a device consisting of a straight channel. Pairs of positive and negative electrodes are placed at fixed intervals along the bottom of the channel as illustrated in Fig. 4.11. The distance  $r$  between the positive and negative electrodes is small compared to the distance  $\ell$  between the pairs. When the electrodes are turned on, the negatively charged DNA molecules move away from the negative electrodes and collect on the positive ones. What is considered here are the experiments using this device as described in Bader et al. (1999). It is worth noting that such devices are now being used for manipulating nanoparticles, both for separation (Wu et al., 2016) and for moving them along predetermined paths (Skaug et al., 2018).

The experiment consists of a repetition of on/off cycles. At the start of each cycle the DNA molecules are attached to the positive electrodes. As illustrated in Fig. 4.12, there are  $P$  molecules, and at the start of the first cycle they are assumed to



**Fig. 4.11** Positive and negative electrodes are placed in pairs at fixed intervals along a channel



**Fig. 4.12** Concentration of the DNA molecules on the positive electrodes at the beginning of the first (a), second (c), and third (e) cycles. When the electrodes are turned off at  $t = 0$ , the molecules spread out as illustrated in (b) and (d). When the electrodes are turned back on at  $t = T$ , the molecules attach themselves to the closest accessible positive electrode as illustrated in (c) and (e)

be attached to the positive electrode at  $x = 0$ . At  $t = 0$  the electrodes are turned off, and when this happens the molecules start spreading out from  $x = 0$ , as determined by the diffusion equation. Using the point source solution, the concentration of DNA is given as

$$c(x, t) = \frac{P}{2\sqrt{\pi Dt}} e^{-x^2/(4Dt)}. \quad (4.23)$$

At  $t = T$  the electrodes are turned back on. When this happens, all of the molecules between the negative electrodes at  $x = -\ell + r$  and  $x = r$  move back to  $x = 0$ , while those between the electrodes at  $x = r$  and  $x = \ell + r$  move to  $x = \ell$  (see Fig. 4.12c). It is possible that some of molecules move far enough to the left that they get past the electrode at  $x = -r$ . To keep this to a minimum we need  $c(-\ell + r, T)$  to be relatively small, and so it is assumed that  $T$  is chosen so that  $4DT \ll (-\ell + r)^2$ . This also guarantees that very few molecules get past the negative electrode at  $x = \ell + r$ . Assuming this is the case, then the number of molecules that end up at  $x = \ell$  is equal to the area under the curve to the right of the red line shown in Fig. 4.12b. From (4.23), it follows that the number is  $\alpha P$ , where

$$\alpha = \frac{1}{2\sqrt{\pi DT}} \int_r^\infty e^{-x^2/(4DT)} dx. \quad (4.24)$$

Given that the total number is  $P$ , then the number that move back to  $x = 0$  is  $(1 - \alpha)P$ . This ends the first cycle.

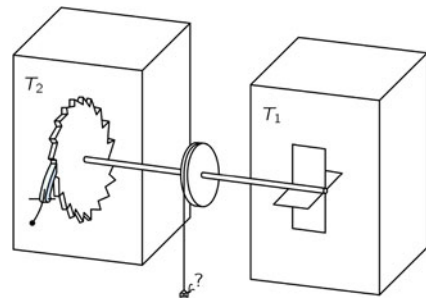
At the start of the second cycle, the molecules are attached to the positive electrodes at  $x = 0$  and  $x = \ell$ . The time variable is reset, and so at  $t = 0$  the electrodes are turned off, and then turned back on at  $t = T$ . Depending on where they are located, the molecules will return to the positive electrodes at  $x = 0$ ,  $x = \ell$ , or  $x = 2\ell$ . Because the same time interval  $T$  has been used, the number that return to  $x = 0$ ,  $x = \ell$ , and  $x = 2\ell$  will be  $(1 - \alpha)^2 P$ ,  $2\alpha(1 - \alpha)P$ , and  $\alpha^2 P$ , respectively. In this way, the off-on cycles use diffusion to move the molecules to the right.

In the experiments in Bader et al. (1999),  $D = 3.24 \times 10^{-12} \text{ m}^2/\text{s}$ ,  $\ell = 20 \text{ }\mu\text{m}$ , and  $T = 1 \text{ s}$ . For these values,  $\alpha \approx 0.216$ . This means that approximately 19 off-on cycles are required to be able to have only 1% of the molecules left at  $x = 0$ , all the rest having moved to one of the electrodes along the positive  $x$ -axis. ■

The history of Brownian ratchets is very interesting. A landmark event was Feynman's description in 1963 of a ratchet and pawl device to lift a bug, as shown in Fig. 4.13. Box  $T_1$  is filled with a gas and contains an axle with vanes attached. The collisions of gas molecules on the vanes generate random forces on the vanes. The ratchet on the other end of the axle, located in box  $T_2$ , only allows rotation in one direction, and this means the wheel rotates in one direction and in the process lifts the bug. This device generated a lot of discussion as it appears to obtain work for free, in other words it seems to be a perpetual motion machine. This is impossible because this would violate the Second Law of Thermodynamics. Nevertheless, there have been numerous attempts to build such a device, or something resembling it. A recent example is an optical trap system involving colloidal particles (Bang et al., 2018), but the efficacy of such a device is still to be determined.

Considerable research has been invested in using ratchets. The underlying idea is that there is a nondirectional source of energy, like heat, along with an externally imposed force that introduces a directionality in the motion. In the case of when the force is turned off and on, as in the DNA experiment considered above, it is called a

**Fig. 4.13** Feynman's ratchet and pawl system for lifting bugs (Feynman et al. 2005)



flashing ratchet. Those interested in learning more about this should consult Hänggi and Marchesoni (2009); Hoffmann (2016), or Lau et al. (2017).

### 4.4.2 A Step Function Initial Condition

The problem to be considered consists of the diffusion equation (4.13), along with the initial condition

$$u(x, 0) = \begin{cases} u_L & \text{if } x < 0, \\ u_R & \text{if } 0 < x. \end{cases} \quad (4.25)$$

This problem can be solved using a slightly modified version of the similarity argument used in Sect. 1.4. The difference is that the dimensionally reduced form in (1.54) is now  $u = u_L F(\eta, z)$ , where  $\eta = x/\sqrt{Dt}$  and  $z = u_R/u_L$ . Carrying out the rest of the similarity analysis one finds that the solution of the diffusion problem is, for  $t > 0$ ,

$$u(x, t) = u_R + \frac{1}{2}(u_L - u_R)\text{erfc}(\eta/2), \quad (4.26)$$

where  $\text{erfc}$  is the complementary error function, which is defined in (1.62).

There is a technical point to be made about the jump in the initial condition. As you will notice, the value at the jump is not specified. The reason is that there is not a consistent, or well-defined, value for the solution at such points. For example, with the above solution, by following the lines  $x = \alpha t$  into the origin of the  $x, t$ -plane, one can get any value of  $u$  between  $u_L$  and  $u_R$ .

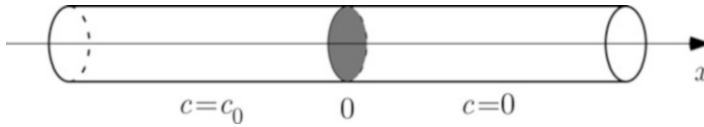
*Example (Determining  $D$ )* An experimental method used to study diffusion involves compartments, and a typical example is shown in Fig. 4.14. The tube is filled with water, and is separated into two compartments. The water on the left, where  $x < 0$ , contains salt with a constant concentration  $c_0$ . The compartment on the right, where  $x > 0$ , contains pure water. At  $t = 0$  the divider separating these two compartments is removed, and this allows the salt to move into the region  $x \geq 0$ . It is assumed that the tube is very long, so the interval can be taken to be  $-\infty < x < \infty$ . With this, assuming the motion is governed by diffusion, then the concentration  $c(x, t)$  of salt along the tube satisfies

$$c_t = Dc_{xx}, \quad \text{for } \begin{cases} -\infty < x < \infty, \\ 0 < t, \end{cases}$$

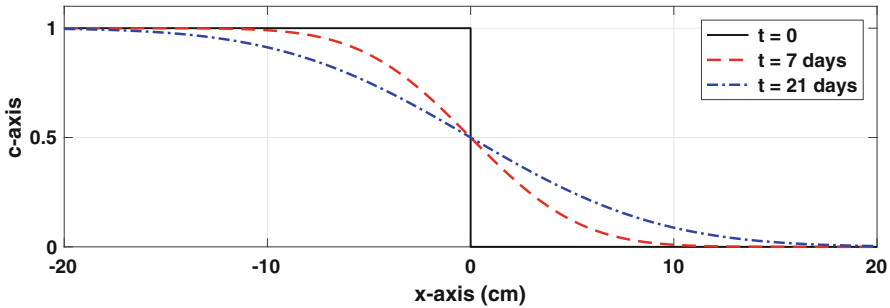
with the initial condition

$$c(x, 0) = \begin{cases} c_0 & \text{if } x < 0, \\ 0 & \text{if } 0 < x. \end{cases}$$





**Fig. 4.14** At the start of the experiment, the water in the left half of the tube contains salt, with concentration  $c_0$ , and the right half contains pure water. The separator between these two regions is removed at  $t = 0$ , and the movement of salt into the right side is then recorded



**Fig. 4.15** The solution (4.27) of the salt diffusion problem at three values of time. In the calculations,  $D = 1.5 \times 10^{-9} \text{ m}^2/\text{s}$  and  $c_0 = 1$

From (4.51), the solution of this problem is

$$c(x, t) = \frac{1}{2} c_0 \operatorname{erfc}(\eta), \quad (4.27)$$

where  $\eta = x/(2\sqrt{Dt})$ .

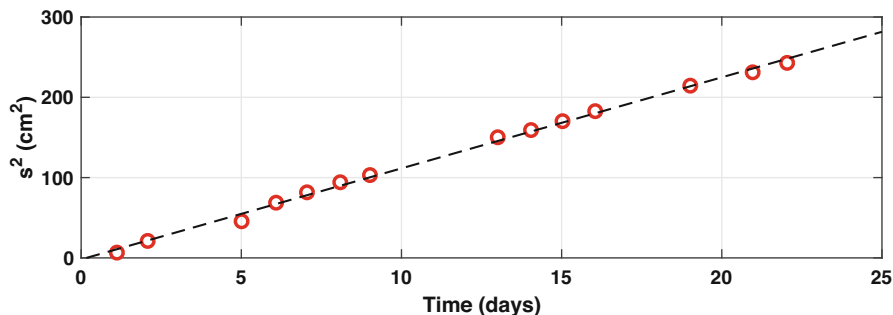
This experiment was used by Booth et al. (1978) to investigate the diffusion of salt in water, and they found that  $D = 1.5 \times 10^{-9} \text{ m}^2/\text{s}$ . Using this value for the diffusion coefficient, the resulting solution (4.27) is shown in Fig. 4.15. This shows that the salt does move into the region on the right. However, what is not at all clear is whether the motion is governed by diffusion, or some other transport mechanism. To check on this we need more specific information obtained from the experiment.

The apparatus they used was able to track the position where  $c$  has a specified value. For example, if they were interested in where  $c = \frac{1}{10} c_0$ , then their device could follow the  $x$  position where this happened. To use this with our solution in (4.27), let  $\bar{c}$  be the specified concentration, and let  $\bar{x}$  be the location where  $c$  has this value. From (4.27),  $2\bar{c} = c_0 \operatorname{erfc}(\bar{x}/(2\sqrt{Dt}))$ . Solving this for  $\bar{x}$  yields

$$\bar{x} = 2\alpha\sqrt{Dt}, \quad (4.28)$$

where

$$\alpha = \operatorname{erfc}^{-1}\left(\frac{2\bar{c}}{c_0}\right). \quad (4.29)$$



**Fig. 4.16** The measured values of  $s^2$ , as defined in (4.30), for the diffusion of salt in water (Booth et al. 1978). The dashed line is a least squares fit to the data points

In their experiments, they followed the positions  $\bar{x}_1$  and  $\bar{x}_2$  for two concentrations,  $\bar{c}_1$  and  $\bar{c}_2$ , and then calculated the distance  $s = \bar{x}_1 - \bar{x}_2$  between these two locations. According to the model, as given in (4.28),

$$s^2 = 4Dt(\alpha_1 - \alpha_2)^2, \quad (4.30)$$

where  $\alpha_1$  and  $\alpha_2$  are the respective values of  $\alpha$  for  $\bar{c}_1$  and  $\bar{c}_2$ . Therefore, the model predicts that the distance squared is linear in time. This is a very strong statement, but does this actually happen? Well, their experimental results are shown in Fig. 4.16 and evidently it does. This is compelling evidence that the diffusion model applies to this system. Moreover, the model shows that the slope of this line can be used to find  $D$ . Just in case you are curious, the value computed using this data is  $D = 1.5 \times 10^{-9} \text{ m}^2/\text{s}$ . ■

Something that is easy to miss in Fig. 4.16 is that  $s = 15 \text{ cm}$  when  $t = 21$  days. In other words, it takes about *three weeks* for the salt to diffuse just 15 cm! Not what you would call a fast mover. Also, the value of the diffusion coefficient is typical for solutes in water. What this means is that diffusion tends to be important over short distances and short time intervals. An indication of this comes from the diffusion coefficient by noting that the value  $D = 10^{-9} \text{ m}^2/\text{s}$  can also be expressed as  $D = 1 \mu\text{m}^2/\text{ms}$ , the implication being that diffusion is significant over distances measured in microns and time measured in milliseconds. This is why diffusion plays such an important role in biological applications related to the function of cells. Movement over larger distances tends to be dominated by convection, which occurs when the fluid flows and in the process carries the molecules with it. The situation is a bit different for diffusion in a gas where the diffusion coefficient is larger, typically by a factor of  $10^4$ . For example, as found in Sect. 4.3.1, for  $\text{O}_2$  in air,  $D = 2 \times 10^{-5} \text{ m}^2/\text{s}$ . This means that diffusion plays an important role over somewhat larger spatial and temporal intervals in a gas. Even so, convection is essential to the movement in a gas, and how to model this transport mechanism will be explored in the next chapters.

## 4.5 Fourier Transform

The random walk considered earlier had no spatial boundaries, the implication being that  $-\infty < x < \infty$ . The usual approach for finding the solution in such situations is to try a transform method. There are many to pick from, and we will consider one of the more well known, the Fourier transform. It possesses one of the distinguishing characteristics of most transforms, and that is that it converts differentiation into multiplication. Exactly what this comment means will be explained below.

We are interested in an unbounded interval, and the specific problem is

$$u_t = Du_{xx}, \quad \text{for } \begin{cases} -\infty < x < \infty, \\ 0 < t, \end{cases} \quad (4.31)$$

with the initial condition

$$u(x, 0) = f(x). \quad (4.32)$$

It is assumed that  $f(x)$  is piecewise continuous with  $\lim_{x \rightarrow \pm\infty} f(x) = 0$ .

To solve the above diffusion problem we introduce the Fourier transform of  $u(x, t)$ , defined as

$$U(k, t) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u(x, t) e^{-ikx} dx. \quad (4.33)$$

Occasionally it is convenient to express the above integral in operator form and write  $U = \mathcal{F}(u)$ . The Fourier transform can be inverted and the formula, in the case that  $u$  is continuous at  $x$ , is

$$u(x, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} U(k, t) e^{ikx} dk. \quad (4.34)$$

In operator form this is written as  $u = \mathcal{F}^{-1}(U)$ . It should be noted that if  $u$  has a jump discontinuity at  $x$ , then the integral in (4.34) does not equal  $u(x, t)$ , but is equal to the average of the jump in  $u$ . Therefore, at a jump discontinuity

$$\frac{1}{2} [u(x^+, t) + u(x^-, t)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} U(k, t) e^{ikx} dk, \quad (4.35)$$

where  $u(x^+, t)$  is the limit from the right, and  $u(x^-, t)$  is the limit from the left.

Given the improper integral in (4.33), it is evident that the definition of  $\mathcal{F}(u)$  requires  $u$  to be reasonably smooth and behave appropriately as  $x \rightarrow \pm\infty$ . For example, the Fourier transform exists if  $\int_{-\infty}^{\infty} |u| dx$  is finite and  $u$  is piecewise continuous. What is not obvious is how the integral of  $U$  in (4.34) produces the original function  $u$ . An argument similar to the one originally employed by Fourier is given in Appendix B. A more formal proof can be found in Weinberger (1995).

One observation that comes from the proof is that it is possible to extend the definition of the Fourier transform and include certain functions that do not go to zero as  $x \rightarrow \pm\infty$ . An example is the periodic function  $f(x) = \cos(\omega x)$ . This requires the introduction of what are known as generalized functions, or distributions. It is not necessary to introduce these for the applications considered here, but those interested in this should consult Friedman (2005).

*Example 1* For the function

$$f(x) = \begin{cases} \alpha & \text{if } a \leq x \leq b, \\ 0 & \text{otherwise,} \end{cases} \quad (4.36)$$

the Fourier transform is

$$\begin{aligned} F(k) &= \frac{1}{\sqrt{2\pi}} \int_a^b \alpha e^{-ikx} dx \\ &= \frac{1}{\sqrt{2\pi}} \frac{i\alpha}{k} (e^{-ikb} - e^{-ika}). \end{aligned} \quad (4.37)$$

This result appears as Property 20 in Table 4.1. ■

*Example 2* For the function  $f(x) = e^{-\alpha|x|}$ , where  $\alpha > 0$ , the Fourier transform is

$$\begin{aligned} F(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ikx - \alpha|x|} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 e^{-(ik - \alpha)x} dx + \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-(ik + \alpha)x} dx \\ &= \sqrt{\frac{2}{\pi}} \frac{\alpha}{\alpha^2 + k^2}. \end{aligned}$$

This result appears as Property 9 in Table 4.1. ■

### 4.5.1 Transformation of Derivatives

The reason the Fourier transform will enable us to solve the diffusion equation is that it converts differentiation into multiplication. To explain what this means we use integration by parts to obtain the following result:

$$\begin{aligned} \mathcal{F}(u_x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u_x e^{-ikx} dx = \frac{1}{\sqrt{2\pi}} \left( u e^{-ikx} \Big|_{x=-\infty}^{\infty} + ik \int_{-\infty}^{\infty} u e^{-ikx} dx \right) \\ &= ik \mathcal{F}(u). \end{aligned} \quad (4.38)$$

**Table 4.1** Inverse Fourier transforms

	$F(k)$	$f(x)$
1.	$F(k)G(k)$	$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(s)g(x-s)ds$
2.	$aF(k) + bG(k)$	$af(x) + bg(x)$
3.	$e^{-iak} F(k)$	$f(x-a)$
4.	$F(k-a)$	$f(x)e^{iax}$
5.	$F(k+a) + F(k-a)$	$2f(x) \cos(ax)$
6.	$F(k+a) - F(k-a)$	$-2if(x) \sin(ax)$
7.	$(ik)^n F(k)$	$\frac{d^n f}{dx^n}$
8.	$\frac{d^n F}{dk^n}$	$(-ix)^n f(x)$
9.	$\frac{1}{a^2+k^2}$	$\frac{1}{a} \sqrt{\frac{\pi}{2}} e^{-a x }$ for $a > 0$
10.	$\frac{k}{a^2+k^2}$	$i \sqrt{\frac{\pi}{2}} e^{-a x } [I_{(0,\infty)}(x) - I_{(-\infty,0)}(x)]$ for $a > 0$
11.	$\frac{\sin(ak)}{k}$	$\sqrt{\frac{\pi}{2}} I_{(-a,a)}(x)$ for $a > 0$
12.	$\frac{1}{\sqrt{a^2-k^2}} I_{(-a,a)}(k)$	$\sqrt{\frac{\pi}{2}} J_0(ax)$ for $a > 0$
13.	$\frac{1}{a+ik}$	$\sqrt{2\pi} e^{-ax} I_{(0,\infty)}(x)$ for $a > 0$
14.	$\frac{1}{(a+ik)^{n+1}}$	$\frac{1}{n!} \sqrt{2\pi} x^n e^{-ax} I_{(0,\infty)}(x)$ for $a > 0$
15.	$\frac{1}{a-ik}$	$\sqrt{2\pi} e^{ax} I_{(-\infty,0)}(x)$ for $a > 0$
16.	$\frac{1}{(a-ik)^{n+1}}$	$\frac{1}{n!} \sqrt{2\pi} (-x)^n e^{ax} I_{(-\infty,0)}(x)$ for $a > 0$
17.	$e^{-a k }$	$\sqrt{\frac{2}{\pi}} \frac{a}{a^2+x^2}$ for $a > 0$
18.	$ke^{-a k }$	$\sqrt{\frac{2}{\pi}} \frac{2iax}{(a^2+x^2)^2}$ for $a > 0$
19.	$e^{-ak^2-ibk}$	$\frac{1}{\sqrt{2a}} e^{-(x-b)^2/(4a)}$ for $a > 0$
20.	$\frac{1}{k} (e^{-ibk} - e^{-iak})$	$-i\sqrt{2\pi} I_{(a,b)}(x)$ for $a < b$
21.	$\frac{\sin^2(ak/2)}{k^2}$	$\frac{1}{2} \sqrt{\frac{\pi}{2}} (a -  x ) I_{(-a,a)}(x)$ for $a > 0$

The indicator function  $I_{(a,b)}(x)$  is defined in (4.41). The general formulas 2.-8. must be modified at a jump discontinuity, as given in (4.35). Also, the numbers  $a$  and  $b$  in this table are real-valued

It has been assumed here that  $u \rightarrow 0$  as  $x \rightarrow \pm\infty$ . In a similar fashion, assuming that  $u_x \rightarrow 0$  as  $x \rightarrow \pm\infty$ , one finds that

$$\mathcal{F}(u_{xx}) = (ik)^2 \mathcal{F}(u). \quad (4.39)$$

The generalization of this to higher derivatives is given in Table 4.1. Therefore, using the Fourier transform, differentiation is transformed into multiplication by  $ik$ .

### 4.5.2 Convolution Theorem

A few of the more well-known formulas for the inverse transform are given in Table 4.1. This includes some of its general properties, which are the first eight entries. These are all derivable directly from the definition of the transform and the properties of integrals. For example, the first one, which is known as the convolution theorem, states that

$$\mathcal{F}\left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(s)g(x-s)ds\right) = F(k)G(k).$$

To prove this, the left-hand side of the above equation is

$$\begin{aligned} & \mathcal{F}\left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(s)g(x-s)ds\right) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(s)g(x-s)e^{-ikx} ds dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(s) \left[ \int_{-\infty}^{\infty} g(x-s)e^{-ikx} dx \right] ds \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(s) \left[ \int_{-\infty}^{\infty} g(z)e^{-ik(z+s)} dz \right] ds \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(s)e^{-iks} ds \left[ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(z)e^{-ikz} dz \right] \\ &= F(k)G(k). \end{aligned}$$

In the above derivation, it is assumed that  $f(x)$  and  $g(x)$  decay fast enough as  $x \rightarrow \pm\infty$  that the improper integrals can be interchanged.

*Example 1* Suppose

$$F(k) = \frac{\sin(3k)}{7k} - 2e^{-4|k|}.$$

To determine the original function  $f(x)$ , we use Property 2, Property 11 with  $a = 3$ , and Property 17 with  $a = 4$ . These are used as follows:

$$\begin{aligned} f(x) &= \mathcal{F}^{-1}\left(\frac{\sin(3k)}{7k} - 2e^{-4|k|}\right) \\ &= \frac{1}{7}\mathcal{F}^{-1}\left(\frac{\sin(3k)}{k}\right) - 2\mathcal{F}^{-1}\left(e^{-4|k|}\right) \\ &= \frac{1}{7}\sqrt{\frac{\pi}{2}} I_{(-3,3)}(x) - \sqrt{\frac{2}{\pi}} \frac{8}{16+x^2}, \end{aligned} \tag{4.40}$$

where  $I_{(a,b)}(x)$  is the *indicator function* and it is defined as

$$I_{(a,b)}(x) \equiv \begin{cases} 1 & \text{if } a < x < b, \\ \frac{1}{2} & \text{if } x = a, b, \\ 0 & \text{otherwise.} \end{cases} \quad (4.41)$$

Introducing the definition of  $I$  into (4.40), then

$$f(x) = \begin{cases} -\sqrt{\frac{2}{\pi}} \frac{8}{16+x^2} & \text{if } 3 < |x|, \\ \frac{1}{7}\sqrt{\frac{\pi}{2}} - \sqrt{\frac{2}{\pi}} \frac{8}{16+x^2} & \text{if } -3 < x < 3, \\ \frac{1}{14}\sqrt{\frac{\pi}{2}} - \sqrt{\frac{2}{\pi}} \frac{8}{16+x^2} & \text{if } x = \pm 3. \end{cases} \quad \blacksquare$$

*Example 2* Suppose

$$F(k) = \frac{1}{2+ik} e^{-3k^2}.$$

This transform is not listed in Table 4.1, however, it is a product of two that are listed. Using Property 13 with  $a = 2$ , the inverse of  $1/(2+ik)$  is  $\sqrt{2\pi} e^{-2x} I_{(0,\infty)}(x)$ . Similarly, using Property 19 with  $a = 3$  and  $b = 0$ , the inverse of  $e^{-3k^2}$  is  $e^{-x^2/12}/\sqrt{6}$ . Therefore, from Property 1 we obtain

$$\begin{aligned} f(x) &= \mathcal{F}^{-1} \left( \frac{1}{2+ik} e^{-3k^2} \right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sqrt{2\pi} e^{-2s} I_{(0,\infty)}(s) \frac{1}{\sqrt{6}} e^{-(x-s)^2/12} ds \\ &= \frac{1}{\sqrt{6}} \int_0^{\infty} e^{-2s-(x-s)^2/12} ds. \end{aligned} \quad (4.42)$$

It is not possible to express the integral in terms of elementary functions, so the above expression is the final answer.  $\blacksquare$

A comment is in order about the Fourier transform and functions with jump discontinuities. As illustrated in (4.36), the transform of a function with a jump is not an issue. The inverse transform, however, is a different matter. Specifically, the inverse of the transform in (4.37) produces the original function  $f(x)$  except at the jump points. At those two points the inverse equals the average of the jump. This means that at  $x = a$ , and at  $x = b$ , the inverse transform equals  $\frac{1}{2}\alpha$ . This is why the indicator function  $I_{(a,b)}$  in (4.41) is defined the way it is at the jump points  $x = a$  and  $x = b$ .

### 4.5.3 Solving the Diffusion Equation

The Fourier transform will enable us to solve the diffusion equation but this brings up a dilemma common in applied mathematics. To use the transform we need to know if the solution satisfies the conditions needed to guarantee that the improper integral in (4.33) is defined. However, we do not know the solution and are therefore not able to check that the conditions are satisfied. What this means is that we will use the transform in a heuristic manner and assume that the transform can be used. Afterwards, once an answer is derived, it is possible to verify directly that it does indeed satisfy the original problem.

To use the Fourier transform to solve the diffusion equation we first take the transform of the equation and obtain

$$\mathcal{F}(u_t) = \mathcal{F}(Du_{xx}).$$

Because the transform is in  $x$  and not  $t$ , then  $\mathcal{F}(u_t) = \frac{d}{dt}\mathcal{F}(u) = U_t$ . With this, and using (4.39), we have that

$$U_t = -Dk^2U. \quad (4.43)$$

We also need to transform the initial condition (4.32), and this gives us

$$U(k, 0) = F(k), \quad (4.44)$$

where  $F(k)$  is the Fourier transform of  $f$ . Solving (4.43), and using (4.44), yields

$$U(k, t) = F(k)e^{-Dk^2t}. \quad (4.45)$$

We now come to the step of trying to determine  $u(x, t)$  given that we know its transform  $U(k, t)$ . One possibility is to determine this from scratch, which means using the definition of the inverse transform in (4.34) and working out the resulting integrals. The specifics of this are outlined in Exercise 4.20. The more conventional approach is to use a table of inverse Fourier transforms, and simply look up the needed formula. Our transform (4.45) is not listed in Table 4.1. However, the formula for  $U$  can be factored as a product  $U = FG$ , and this will enable us to find the inverse. Setting  $G = e^{-Dk^2t}$ , then, from Table 4.1,

$$g(x) = \frac{1}{\sqrt{2a}}e^{-x^2/(4a)}, \quad (4.46)$$

where  $a = Dt$ . With this, and the convolution property, we obtain

$$\begin{aligned} u(x, t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(s)g(x-s)ds \\ &= \frac{1}{2\sqrt{\pi Dt}} \int_{-\infty}^{\infty} f(s)e^{-(x-s)^2/(4Dt)}ds. \end{aligned} \quad (4.47)$$



This is the sought after solution of the diffusion problem. What is interesting is that it consists of the integral of the given initial condition multiplied by the point source solution in (4.18). For those who might question some of the steps used to obtain this result, it is a simple matter to show that (4.47) does indeed satisfy the diffusion equation. What is not as straightforward is verifying that (4.47) satisfies the initial condition (4.32). Taking the limit  $t \rightarrow 0^+$  requires some careful analysis of what happens in the neighborhood of  $s = x$  and a proof can be found in Mikhlin (1970).

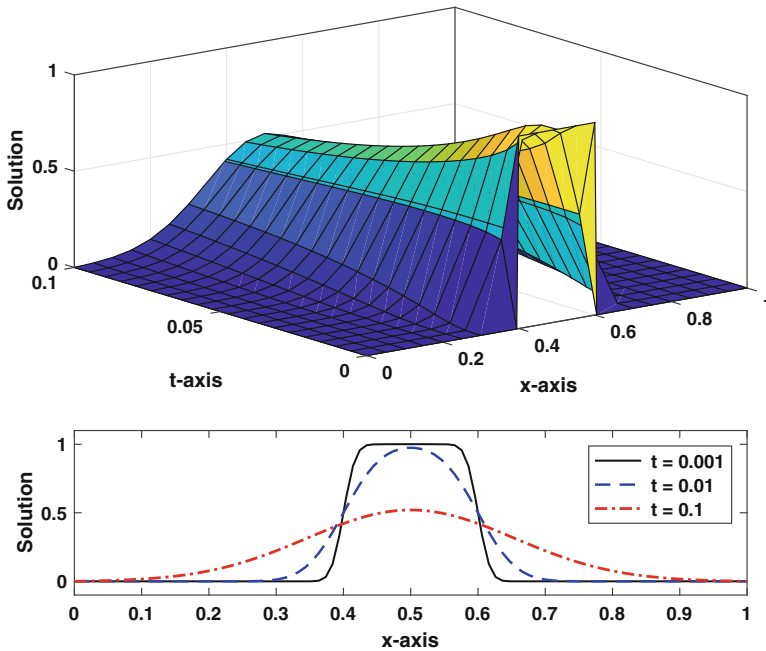
*Example 1* If the initial condition is

$$f(x) = \begin{cases} 1 & \text{if } a < x < b, \\ 0 & \text{otherwise,} \end{cases} \quad (4.48)$$

then, from (4.47), the solution is

$$u(x, t) = \frac{1}{2\sqrt{\pi Dt}} \int_a^b e^{-(x-s)^2/(4Dt)} ds. \quad (4.49)$$

This function is shown in Fig. 4.17, both as time slices and the solution surface for  $0 \leq t \leq 0.1$ . This illustrates several of the characteristic properties of a solution of the diffusion equation. One is that even with an initial condition that contains



**Fig. 4.17** Solution (4.49) of the diffusion equation when  $f(x)$  is given in (4.48), with  $a = 0.4$ ,  $b = 0.6$ , and  $D = 0.1$ . Shown is the solution surface as well as the solution profiles at specific time values

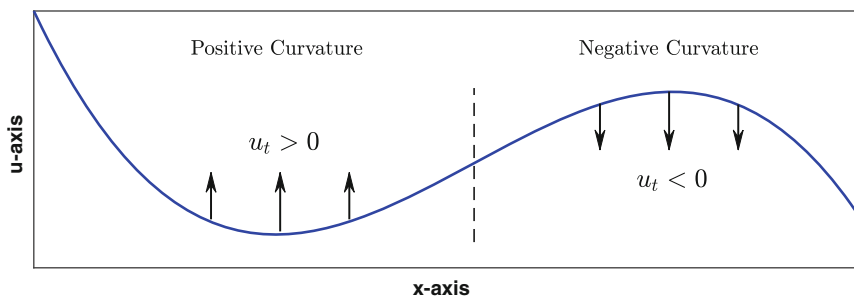
jumps, the solution for  $0 < t$  is smooth. A second observation is that because the exponential function is positive, then the solution in (4.49) is never zero for  $t > 0$ . This means, for example, that even though the solution starts out as zero at  $x = 1000$ , and the nonzero portion of  $f(x)$  is far away, that the solution is nonzero at this position for any value of  $t > 0$ . This means diffusion occurs with infinite speed, which is physically unrealistic. The usual argument made is that the solution decays rapidly as  $x \rightarrow \pm\infty$ , so the consequence of this is minimal. The fact is that the diffusion equation is a mathematical model, and as such it is an approximation, albeit an effective approximation. Nevertheless, it is interesting that a random walk with steps of finite speed can give rise to macroscopic motion of infinite speed. This paradox has been the subject of numerous studies, two of the more recent being Keller (2004) and Aranovich and Donohue (2007). ■

How the three curves shown in Fig. 4.17 change with  $t$  can be explained using the terms appearing in the diffusion equation. Recall from calculus that a curve  $y = f(x)$  is concave up if  $f'' > 0$ , and it is concave down if  $f'' < 0$ . So, if  $u$  satisfies  $Du_{xx} = u_t$ , and if  $u$  is concave up in  $x$ , so  $u_{xx} > 0$ , then  $u_t > 0$ , and this means that  $u$  is increasing in  $t$ . The opposite happens if  $u$  is concave down in  $x$ . A schematic of the two possibilities is given in Fig. 4.18. With this observation, given the concave down nature of  $u$  in Fig. 4.17, the decrease in the solution is expected.

*Example 2* Suppose the initial condition is

$$f(x) = \begin{cases} u_1 & \text{if } x < 0, \\ u_2 & \text{if } 0 < x. \end{cases} \quad (4.50)$$

It would seem, at first sight, that this is very similar to the previous example. What is wrong with this observation is that the above function does not satisfy the requirement that  $\lim_{x \rightarrow \pm\infty} f(x) = 0$ . Because of this, the Fourier transform of  $f(x)$  does not exist, and so the method used to derive the solution (4.47) is not valid for this problem. However, all is not lost. Irrespective of how it was derived, as long as  $f(x)$  is well behaved, (4.47) is the solution of the problem.



**Fig. 4.18** The change in the solution of the diffusion equation is determined by the local curvature in the solution. The result is the eventual straightening of the curve

The reason is that (4.47) satisfies the diffusion equation and the given initial condition. This happens, for example, if  $f(x)$  is bounded and piecewise continuous. The function in (4.50) satisfies these conditions, and therefore the solution of the resulting diffusion problem is

$$\begin{aligned} u(x, t) &= \frac{u_1}{2\sqrt{\pi Dt}} \int_{-\infty}^0 e^{-\frac{(x-s)^2}{4Dt}} ds + \frac{u_2}{2\sqrt{\pi Dt}} \int_0^{\infty} e^{-\frac{(x-s)^2}{4Dt}} ds \\ &= \frac{u_1}{2\sqrt{\pi Dt}} \int_0^{\infty} e^{-\frac{(x+r)^2}{4Dt}} dr + \frac{u_2}{2\sqrt{\pi Dt}} \int_0^{\infty} e^{-\frac{(x-s)^2}{4Dt}} ds \\ &= \frac{u_1}{\sqrt{\pi}} \int_{\eta/2}^{\infty} e^{-z^2} dz + \frac{u_2}{\sqrt{\pi}} \int_{-\infty}^{\eta/2} e^{-w^2} dw, \end{aligned}$$

where  $\eta = x/\sqrt{Dt}$ . This can be written in terms of the complementary error function  $\operatorname{erfc}(\eta)$ , given in (1.62), as follows:

$$u(x, t) = u_2 + \frac{1}{2}(u_1 - u_2)\operatorname{erfc}(\eta/2). \quad (4.51)$$

As it must, this is the same solution given in (4.26) that we obtained using a similarity variable. ■

The situation occurring in the previous example is not unusual, and it is worth discussing a bit more. To be able to use the Fourier transform, it is necessary to impose rather strict conditions on the functions in the problem. However, the formula that is derived for the solution turns out to be defined for a much broader class of functions than originally assumed. In this case, the formula for the solution becomes the center of attention, and the method that was used to derive the formula is effectively forgotten. This is a very fortunate situation, but the caveat is that care must be taken to make sure that the formula is defined for the functions that are used.

## 4.6 Continuum Formulation of Diffusion

The approach up to this point has been to consider the motion at the micro (or discrete) level and then consider what happens as one passes to the macro (or continuous) level. In this section we will simply start with the continuous description and not concern ourselves with what might, or might not, be happening at the micro level. This is the more conventional method used to derive the diffusion equation.

A good example illustrating the approach is the one used in the original development of the subject by Fick (1885). He noticed that when salt is poured into water the concentration of salt slowly spreads out and eventually becomes uniformly distributed in the water. To obtain an equation for the concentration we will assume

the motion is only along the  $x$ -axis. With this, let  $c(x, t)$  designate the concentration of salt, which in this context has the dimensions of number of particles per unit length. This is sometimes referred to as the linear density.

### 4.6.1 Balance Law

The equation for  $c$  will be derived from a balance law, and to explain how consider an interval  $a \leq x \leq b$ . The number of salt particles in this interval can change for only two reasons. First, they can move along the  $x$ -axis and therefore they can move in or out of the interval. Although we do not know exactly how the salt is moving, it does and therefore let  $J(x, t)$  designate the net number of salt particles that pass  $x$  per unit time. The function  $J$  is known as the *flux*. The second way the number of particles in the interval can change is that they are created or destroyed within the interval. This could happen, for example, through a chemical reaction. For this possibility we introduce the function  $Q(x, t)$ , which gives the number of particles created at  $x$  per unit time. With this, our balance law has the form

$$\frac{d}{dt} \int_a^b c(x, t) dx = J(a, t) - J(b, t) + \int_a^b Q(x, t) dx. \quad (4.52)$$

In words, this equation states that the rate of change in the total number of salt particles in the interval is due to the movement of the particles across the endpoints, this is the  $J(a, t) - J(b, t)$  expression, and to the creation or destruction of the particles within the interval. Using the Fundamental Theorem of Calculus, the above integral can be written as

$$\int_a^b \frac{\partial c}{\partial t} dx = - \int_a^b \frac{\partial J}{\partial x} dx + \int_a^b Q(x, t) dx.$$

This can be rewritten as

$$\int_a^b \left( \frac{\partial c}{\partial t} + \frac{\partial J}{\partial x} - Q \right) dx = 0.$$

This equation holds for any interval. As shown in analysis, if the integral of a continuous function is zero over every interval, then it must be that the function is identically zero. Because of this we conclude that

$$\frac{\partial c}{\partial t} = - \frac{\partial J}{\partial x} + Q. \quad (4.53)$$

This is the balance law we are looking for. In its present form, it is very general and what we need to do is determine, or specify, the functions  $J$  and  $Q$  for the problem we are working on, namely the diffusion of salt in water.

The boundary conditions most often used when solving (4.53) involve either prescribing the value of the concentration at an endpoint, or its flux. As an example, if the interval is  $a < x < b$ , and the concentration is prescribed on the left, and the flux on the right, then the resulting boundary conditions would have the form

$$c(a, t) = c_0, \quad J(b, t) = J_0,$$

where  $c_0$  and  $J_0$  are given.

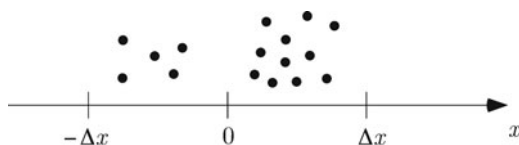
As in the kinetic problems of the last chapter, steady-state solutions play an important role in the analysis of diffusion problems. In this case,  $c$  is a steady-state solution if it is independent of  $t$ , and it satisfies

$$-\frac{\partial J}{\partial x} + Q = 0.$$

For example, if  $J = -Dc_x$ , and  $Q = c^3$ , then the above equation becomes  $Dc_{xx} + c^3 = 0$ . Also, the steady-state solution is required to satisfy the boundary conditions but not the initial condition (this is assuming that the boundary conditions do not involve explicit functions  $t$ ).

### 4.6.2 Fick's Law of Diffusion

The assumption used to specify the flux is that, due to diffusion, the particles in regions of higher concentration will tend to move toward regions of lower concentration. The situation is shown schematically in Fig. 4.19. As shown, there is a small number  $N_\ell$  of particles in the left bin, and a larger number  $N_r$  in the bin on the right. According to the rules of a random walk, in a time step, approximately half of those on the right will move into the bin on the left, and approximately half of those on the left will move to the bin on the right. The flux is the net difference over the time interval, and so  $J = \frac{1}{2}(N_\ell - N_r)/\Delta t$ . To express this using continuum variables, the total number of particles in the interval  $a < x < a + \Delta x$  is



**Fig. 4.19** The flux at  $x = 0$  depends on the difference in the number  $N_\ell$  of particles just to the left of  $x = 0$  and the number  $N_r$  just to the right of  $x = 0$

$$N = \int_a^{a+\Delta x} c(s, t) ds \\ \approx \Delta x c(a, t).$$

With this,

$$N_\ell \approx \Delta x c(-\Delta x, t), \quad \text{and} \quad N_r \approx \Delta x c(0, t).$$

Using Taylor's theorem, we have that the flux is

$$J \approx \frac{1}{2\Delta t} [\Delta x c(-\Delta x, t) - \Delta x c(0, t)] \\ \approx \frac{\Delta x}{2\Delta t} \left[ c(0, t) - \Delta x \frac{\partial c}{\partial x}(0, t) + \cdots - c(0, t) \right] \approx -\frac{\Delta x^2}{2\Delta t} \frac{\partial c}{\partial x}(0, t).$$

The above approximation for the flux at  $x = 0$  provides motivation for the assumption made in the continuum formulation. Specifically, it is assumed that the flux is given as

$$J = -D \frac{\partial c}{\partial x}, \quad (4.54)$$

where  $D$  is a positive constant known as the diffusion coefficient. This is known as *Fick's law of diffusion*, or when applied to temperature distributions it goes by the name of the *Fourier law of heat conduction*.

To complete the derivation, it is assumed that the particles are not created or destroyed. In this case  $Q = 0$  in (4.53), and the balance equation reduces to the diffusion equation

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2}. \quad (4.55)$$

The derivation of this result has required minimal effort because it uses the general balance law along with the constitutive law in (4.54). For this reason, it is favored in most derivations of the diffusion equation.

The formula for the flux given in (4.54) is an example of a constitutive law, and we will come across several of these in the later chapters. Even though we used the random walk to help motivate this assumption, it is important to understand that (4.54) does not assume that the particles are undergoing a random walk. It only assumes that the flux is proportional to the spatial derivative, and what might be going on at the molecular level is not stated.

The more typical method for determining a constitutive law is to measure  $J$  experimentally, and then use this information to specify the function. This approach is used multiple times in the chapters to follow, as evidenced by the data given in Figs. 5.6, 6.5, 9.2, and 9.3. Although this approach is required when using a

continuum model, a data-driven formulation does not explain why the flux depends on the specific variables appearing in the constitutive law. This was the reason for starting this chapter with the random walk analysis, because it illustrates how a microscale movement can be used to explain macroscale motion. An active area of research addresses this issue for more complex problems, attempting to use quantum or molecular theories to derive the appropriate constitutive law. Those who are interested in this can find an introduction to this area in Balluffi et al. (2005) and Lucas (2007).

*Example (Using the Flux to Find  $D$ )* Fluid saturated soil, a sponge filled with water, and articular cartilage are examples of biphasic materials. They are formed from two constituents, a porous solid with the pores filled with water. Given a sample of length  $\ell$ , then the displacement  $u(x, t)$  of such a material is governed by the diffusion equation

$$D \frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t}, \quad \text{for } \begin{cases} 0 < x < \ell, \\ 0 < t. \end{cases} \quad (4.56)$$

In the experiments described in Holmes et al. (1983), the sample is held at  $x = \ell$  and the flux is prescribed at  $x = 0$ . The corresponding boundary conditions are

$$u(\ell, t) = 0, \quad (4.57)$$

and

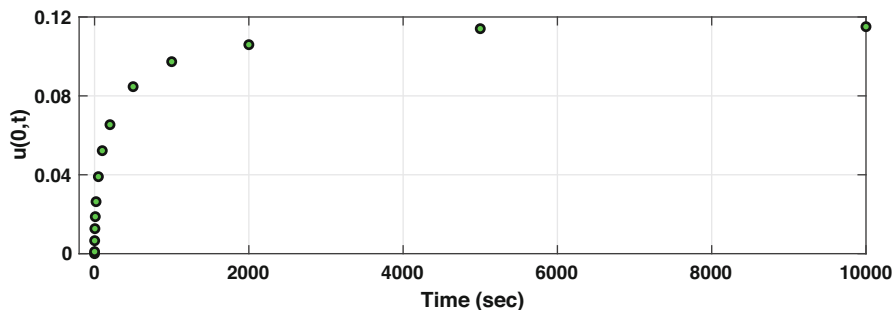
$$D \frac{\partial u}{\partial x}(0, t) = -\alpha. \quad (4.58)$$

Also, the initial condition is

$$u(x, 0) = 0. \quad (4.59)$$

What is measured in the experiments is the value of  $u(0, t)$ , and a typical result is shown in Fig. 4.20. These data are going to be used to address two questions. First, in looking at the values of  $u(0, t)$ , how do you know that the response is governed by the diffusion equation? Second, assuming that the diffusion equation is correct, can you use this information to determine  $D$ ? Answering these questions will involve solving the diffusion problem, but we need to be selective in how this is done. For example, it is possible to find the solution using separation of variables. However, this is not a particularly useful method for this example because it is very difficult to develop an intuitive understanding of the solution from the Fourier series. Instead, we will derive approximations of the solution using what we know about the diffusion equation and the data in Fig. 4.20.

- *Steady State* The first approximation relates to the steady-state solution. It is seen in Fig. 4.20 that  $u(0, t)$  approaches a steady state as  $t \rightarrow \infty$ . From (4.56),



**Fig. 4.20** The values of  $u(0, t)$  measured in response to a prescribed flux. The data are from a test on articular cartilage (Holmes et al. 1983)

the steady state satisfies  $u_{xx} = 0$  along with the boundary conditions in (4.57) and (4.58). The corresponding solution is

$$u = \frac{\alpha}{D}(\ell - x). \quad (4.60)$$

With this, the solution at  $x = 0$  is  $u = \alpha\ell/D$ . Given that  $\alpha$  and  $\ell$  are known, then we can use this equation to find  $D$ . What is needed is the steady-state value of  $u$  at  $x = 0$ . It appears from the data in Fig. 4.20 that the steady state has almost been reached at  $t = 10,000$  s. Using this approximation, then the diffusion coefficient can be calculated using the formula  $D = \alpha\ell/u(0, 10,000)$ .

The steady state has provided the answer to the second question. This leaves the issue of how the data in Fig. 4.20 can be used to help verify that the diffusion equation is the correct model. For this we consider what happens for small values of time.

- *Small Time Approximation* The solution starts out as  $u = 0$ , and what is responsible for causing the solution to be nonzero is the flux boundary condition (4.58). How this information moves through the interval is governed by the diffusion equation, and an indication of what happens can be derived from Fig. 4.17. Namely, it takes a certain amount of time for the nonzero part of the solution to move across the interval and appreciably affect what is happening at  $x = \ell$ . Up until this happens, we can assume that the sample is infinitely long, and replace (4.57) with the condition that

$$u \rightarrow 0 \text{ as } x \rightarrow \infty. \quad (4.61)$$

It is understood that in this approximation the diffusion equation is being solved, not on a finite spatial interval, but for  $0 < x < \infty$ . The easiest way to solve the problem in this case is using similarity variables. The only dimensional quantities appearing in this problem, other than  $u$ , are  $x$ ,  $t$ ,  $D$ , and  $\alpha$ . Therefore, it follows



that  $u = F(x, t, D, \alpha)$ . To reduce this using dimensional analysis note  $[\alpha] = [Du]/L = [u]L/T$ . Using an argument very similar to the one given in Sect. 1.4, it is found that

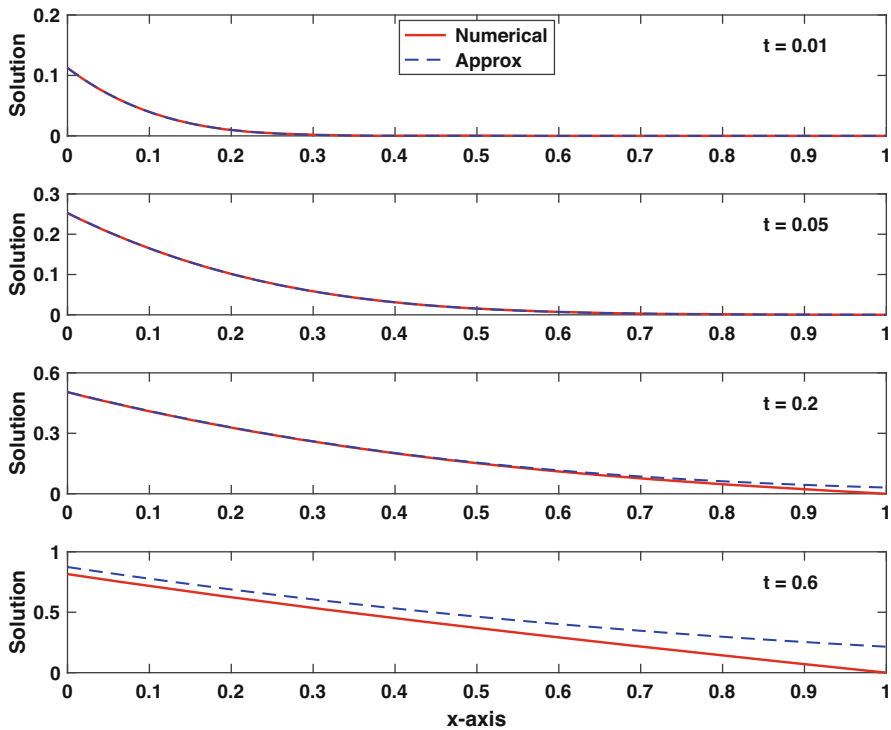
$$u = \alpha \sqrt{\frac{t}{D}} f(\eta), \quad (4.62)$$

where  $\eta = x/\sqrt{Dt}$ . Substituting this into the diffusion equation, and rearranging things a bit, yields

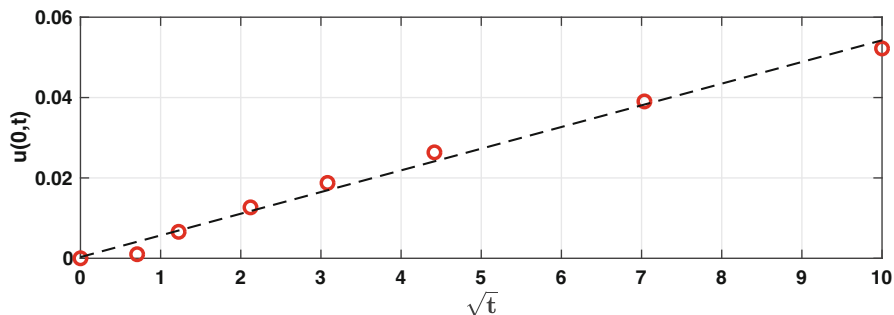
$$f'' + \frac{1}{2}\eta f' - \frac{1}{2}f = 0. \quad (4.63)$$

Staring at this equation for a few moments it is seen that  $f = \eta$  is a solution. This enables us to use reduction of order to find the general solution. This is done by assuming  $f(\eta) = \eta g(\eta)$ , and using (4.63) to find  $g$ . The result is that

$$f(\eta) = a\eta + b\left(e^{-\eta^2/4} - \frac{1}{2}\eta \int_{\eta}^{\infty} e^{-s^2/4} ds\right), \quad (4.64)$$



**Fig. 4.21** Comparison between the numerical solution of the diffusion problem and its approximate solution given in (4.65)



**Fig. 4.22** The data for  $0 \leq t \leq 100$  from Fig. 4.20 plotted as a function of  $\sqrt{t}$ . The data clearly show the linear dependence predicted in (4.66). The dashed line is a least squares fit to the data points

where  $a$  and  $b$  are arbitrary constants. The condition in (4.61), and the initial condition (4.59), require that  $f(\infty) = 0$ . From this it follows that  $a = 0$ . From the flux condition (4.58) one finds that  $f'(0) = -1$ , and this means that  $b = 2/\sqrt{\pi}$ . The resulting solution is

$$\begin{aligned} u(x, t) &= 2\alpha \sqrt{\frac{t}{\pi D}} \left( e^{-\eta^2/4} - \frac{1}{2} \eta \int_{\eta}^{\infty} e^{-s^2/4} ds \right) \\ &= 2\alpha \sqrt{\frac{t}{\pi D}} \left( e^{-\eta^2/4} - \frac{\sqrt{\pi}}{2} \eta \operatorname{erfc}(\eta/2) \right). \end{aligned} \quad (4.65)$$

To illustrate the accuracy of this approximation, it is plotted in Fig. 4.21 along with the numerical solution of the problem. For this comparison,  $\ell = D = \alpha = 1$ . As expected, the two solutions are very close until the disturbance effectively reaches the right endpoint. Once that happens the solution of the diffusion problem rapidly approaches the steady-state solution, which in this case is  $u = 1 - x$ . One of the more important conclusions that comes from (4.65) is that

$$u(0, t) = 2\alpha \sqrt{\frac{t}{\pi D}}. \quad (4.66)$$

This is the result we were looking for because this is what is measured in the experiment. It shows that if the diffusion model is correct, then the deformation of the surface must increase as  $\sqrt{t}$ , at least for small values of  $t$ . It is difficult to see if this happens in Fig. 4.20, so the data are replotted in Fig. 4.22 as a function of  $\sqrt{t}$ , for  $0 \leq t \leq 100$ . Although the linear dependence seen in this figure does not prove that the diffusion model is correct, it is a very compelling evidence that it is. ■

*Example (Drift Diffusion)* Up to this point the particle's motion is due exclusively to Brownian motion, but we now include the contribution from an external force. To determine how this affects the flux, suppose the external force results in the particles having a velocity  $v_d$ , what we will refer to as the drift velocity. The flux in this case can be written as  $J = J_{\text{diff}} + J_{\text{drift}}$ , where  $J_{\text{diff}}$  is given in (4.54). To determine  $J_{\text{drift}}$ , we refer back to Fig. 4.19. Assuming  $v_d$  is positive, then over a small time interval  $\Delta t$ , the particles that are able to cross  $x = 0$  will be within an interval of width  $v_d \Delta t$ . This number is approximately  $c(0, t)v_d \Delta t$ , and the resulting flux, which is the number per time interval, is  $J_{\text{drift}} = c(0, t)v_d$ . This is for  $x = 0$ , and is applicable for both positive and negative  $v_d$ . Generalizing this

$$J_{\text{drift}} = v_d c,$$

and from this we obtain the following constitutive relation for the flux:

$$J = v_d c - D \frac{\partial c}{\partial x}. \quad (4.67)$$

The corresponding equation of motion is

$$c_t = D c_{xx} - v_d c_x. \quad (4.68)$$

This is an example of a convection-diffusion equation, and it can be solved in a straightforward manner using the Fourier transform (see Exercise 4.17). Rather than doing that, we will work out the steady-state solution, which means we find the function that is independent of  $t$  and satisfies  $D c_{xx} - v_d c_x = 0$ . Assuming  $c = e^{rx}$  one finds from the differential equation that  $r = 0, v_d/D$ . From this it follows that the general solution of the steady-state equation, for  $v_d \neq 0$ , is  $c = a + b e^{v_d x/D}$ , where  $a$  and  $b$  are arbitrary constants. ■

A drift velocity can arise for a variety of reasons, and one example is when the particles are charged and an electrical potential is applied. The electric field will induce the particles to move, and there are various ways to determine the resulting formula for  $J_{\text{drift}}$ . One approach simply makes the assumption that the velocity is proportional to the electric field. The corresponding constitutive law for the drift velocity is  $v_d = \mu E$ , where  $E$  is the electric field and  $\mu$  is a constant known as the mobility. It is possible to obtain this result using a more physically based argument, and this is contained in the next example.

*Example (Nernst-Planck Law)* Suppose the charged particles are in solution. In this case, the motion of the particles induced by the electric field will be resisted by the viscosity of the fluid. Assuming that the motion is steady, then the resulting drift velocity will correspond to when these forces balance. The electric field force is  $qE$ , where  $q$  is the charge of a particle. To determine the viscous force, it is assumed that the particle is spherical and its velocity is relatively slow. In this case, from (1.22), the viscous drag force is  $D_F = 6\pi\eta r v_d$ , where  $r$  is the radius of

the particle and  $\eta$  is the dynamic viscosity of the solvent. Using (4.15) this can be rewritten as  $D_F = kT v_d / D$ . Equating the electric and viscous force it follows that  $v_d = qED / (kT)$ . The resulting constitutive law for the flux takes the form

$$J = D \left( -\frac{\partial c}{\partial x} + \frac{qE}{kT} c \right), \quad (4.69)$$

which is known as the Nernst-Planck law. The resulting diffusion equation is given in (4.68), where  $v_d = qED / (kT)$ . The steady-state solution, which produces a zero flux, is

$$c = c_0 e^{xqE/(kT)},$$

where  $c_0$  is a constant. This is an example of what is known as a Boltzmann distribution for the concentration. ■

As a final comment on drift diffusion, although it is routinely arises in applications, there are limitations when (4.68) can be used. An assumption inherent in the derivation of this equation is that the drift velocity is constant, and, therefore, independent of  $c$ . Although this is a reasonable assumption at low concentrations and small drift velocities, it is very questionable once the concentrations and velocities increase. This observation is central to the next chapter, where the relationship between the concentration and velocity plays a central role in the analysis.

### 4.6.3 Reaction-Diffusion Equations

Up to this point we have assumed the particles are not created or destroyed. An example of a situation where this does not happen is when the particles are undergoing chemical reactions. To illustrate what effect this has on the diffusion equation suppose we have two species,  $A$  and  $B$ , and they are undergoing the reversible reaction



Assuming, for the moment, that there is no diffusion, then the resulting kinetic equations are obtained using the law of mass action, and the result is

$$\begin{aligned} \frac{dA}{dt} &= -k_1 A + k_{-1} B, \\ \frac{dB}{dt} &= k_1 A - k_{-1} B. \end{aligned}$$

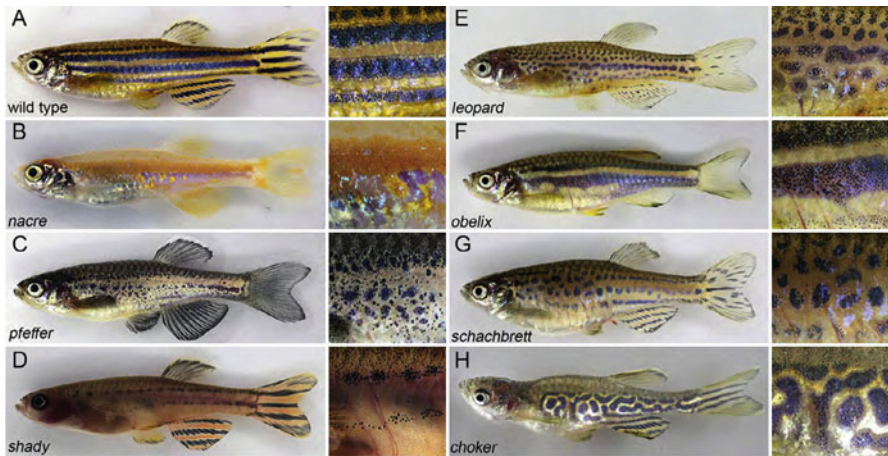
The right-hand sides of these two equations are our source terms. Namely, for species  $A$  we have that  $Q = -k_1 A + k_{-1} B$ , and for species  $B$  we have  $Q = k_1 A - k_{-1} B$ . The resulting diffusion equations are

$$\begin{aligned}\frac{\partial A}{\partial t} &= D_a \frac{\partial^2 A}{\partial x^2} - k_1 A + k_{-1} B, \\ \frac{\partial B}{\partial t} &= D_b \frac{\partial^2 B}{\partial x^2} + k_1 A - k_{-1} B.\end{aligned}$$

It has been assumed here that the diffusion coefficients for the two species are different. The above system of equations is an example of reaction-diffusion equations.

*Example (Pattern Formation)* A well-researched question in developmental biology is how cells in an organism arrange themselves to form patterns. One of the objectives is to understand how the very colorful, and geometrically intricate, patterns seen on butterflies, seashells, birds, etc. are formed. Because the zebrafish is easily manipulated genetically, and there is a library of mutant lines, it has become the focus of study in regard to pattern formation. Some of the possible variations for a zebrafish are shown in Fig. 4.23.

One possible mechanism, first described by Turing (1952), is that two or more chemicals diffuse through an embryo and react with each other until a stable pattern of chemical concentrations is reached. One of the more popular models for this is due to Gray and Scott, in which (4.70) is replaced with  $2A + B \rightarrow 3A$ ,  $A \rightarrow P$ .



**Fig. 4.23** Pigmentation patterns in mutants of zebrafish used in the study of the Turing mechanism (Irion et al., 2016)

It is also assumed there is a constant inflow of  $B$ . The resulting equations are (Gray and Scott, 1994)

$$\begin{aligned}\frac{\partial A}{\partial t} &= D_a \frac{\partial^2 A}{\partial x^2} + k_1 A^2 B - k_2 A, \\ \frac{\partial B}{\partial t} &= D_b \frac{\partial^2 B}{\partial x^2} - k_1 A^2 B + \mu.\end{aligned}$$

More elaborate models have certainly been proposed, but the recent research in this area has concentrated on experimentally identifying the molecular mechanism associated with Turing patterns. Those interested in this topic should consult Meinhardt (2012), Watanabe and Kondo (2015), and Irion et al. (2016). ■

## 4.7 Random Walks and Diffusion in Higher Dimensions

It is interesting to consider how to extend random walks to multiple dimensions. The basic idea is that starting at  $\mathbf{x}_0$  the first step in the walk produces a new position  $\mathbf{x}_1$ . Earlier we assumed the length of a step is fixed, and we will do the same here. If the step length is  $h$ , then the formula connecting  $\mathbf{x}_1$  with  $\mathbf{x}_0$  is

$$\mathbf{x}_1 = \mathbf{x}_0 + h\mathbf{u}_1.$$

In one dimension it is possible to move only left or right, and in this case  $\mathbf{u}_1$  is randomly chosen to be  $\pm 1$ . In higher dimensions,  $\mathbf{u}_1$  is a randomly selected direction, or more precisely, a randomly chosen unit vector. Once we have  $\mathbf{x}_1$ , the next step  $\mathbf{x}_2$  in the walk is calculated in a similar fashion. The only difference is that we randomly select a new direction vector  $\mathbf{u}_2$ . Generalizing this procedure, the resulting formula for the position at time step  $n$  is

$$\mathbf{x}_n = \mathbf{x}_{n-1} + h\mathbf{u}_n, \quad (4.71)$$

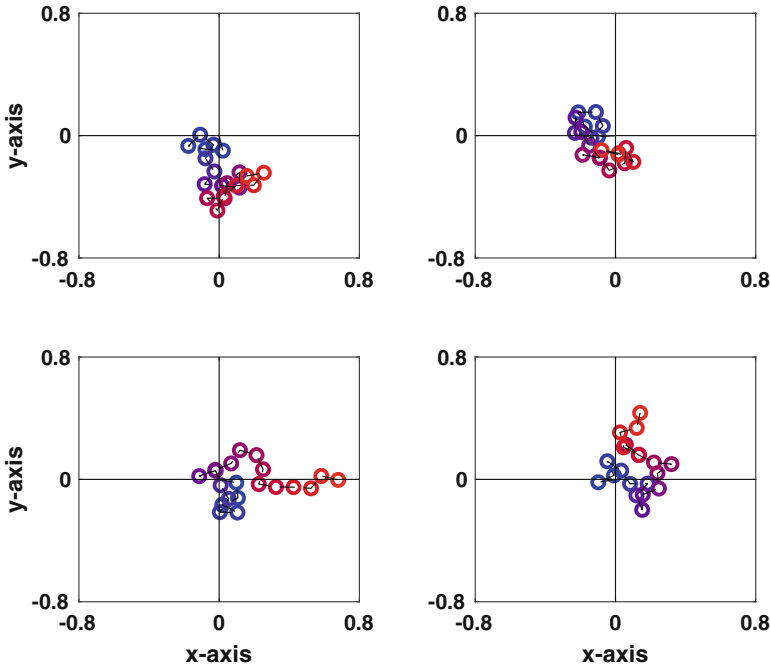
where  $\mathbf{u}_n$  is a randomly chosen unit vector.

To visualize what happens consider two dimensions, where  $\mathbf{x} = (x, y)$ . In this case the direction for  $\mathbf{x}_n$  can be written in terms of the polar coordinate angle  $\theta$ . With this, (4.71) becomes

$$x_n = x_{n-1} + h \cos(\theta_n), \quad (4.72)$$

$$y_n = y_{n-1} + h \sin(\theta_n), \quad (4.73)$$

where  $\theta_n$  is randomly chosen from the interval  $[0, 2\pi)$ . The first 20 positions calculated using this formula are shown in Fig. 4.24 using  $h = 0.1$ . Four random walks are shown, and not unexpectedly they are quite different from one another.



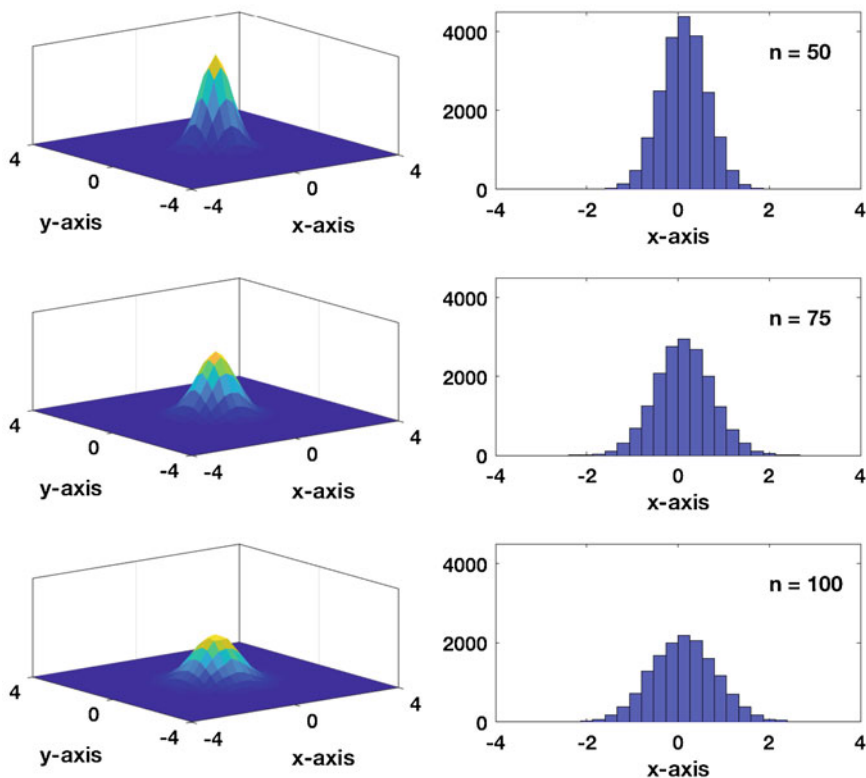
**Fig. 4.24** Random walks in the plane. Each starts at  $x = y = 0$  and has step length  $h = 0.1$ . The color of the particles changes smoothly from blue, when  $n = 1$ , to red, when  $n = 20$

Nevertheless, a few observations can be made. First, none of them has come close to reaching the maximum obtainable distance of  $20h = 2$ . This is not surprising because to reach the maximum distance the same angle would need to be used at each step, and this is highly unlikely. Second, there is a propensity for the positions to be close to the origin.

It is natural to ask if the positions follow the Gaussian distribution found for the one-dimensional random walks shown in Fig. 4.6. To determine this numerically, suppose we take 100,000 random walks, all starting at  $x = y = 0$ , and look at the distribution of positions at various time points. The results are shown in Fig. 4.25 when  $n = 50$ ,  $n = 75$ , and  $n = 100$ . It appears that at the beginning the particles are closer to the origin but as time increases they move outward. These appear to be a two-dimensional version of the point source solution (4.18) we derived earlier. They are, and it is possible to show that the multidimensional version of the point source solution is

$$\frac{1}{(4\pi Dt)^{d/2}} e^{-r^2/(4Dt)}, \quad (4.74)$$

where  $d$  equals the number of spatial dimensions and  $r$  is the radial distance to the origin. For the distributions in Fig. 4.25,  $d = 2$  and this means that the amplitude



**Fig. 4.25** Distribution of positions at  $n = 50, 75, 100$  for random walks using 100,000 particles. On the left are the distributions in the plane, and on the right is the distribution along the  $x$ -axis

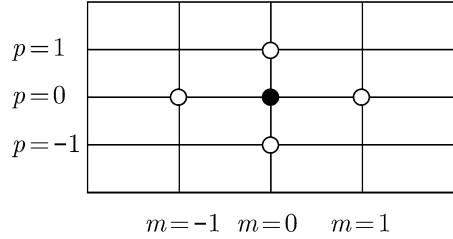
decreases as  $1/t$ . This dependence is evident in the distributions, as doubling the value of  $n$  results in a reduction in the amplitude by approximately a factor of two. This leaves open the question of how to derive (4.74), and this will be done later, for  $d = 2$ , after the diffusion equation has been derived.

### 4.7.1 Diffusion Equation

It is enough to derive the diffusion equation for two spatial dimensions. Also, we will limit the directions used in the random walk. Specifically, the angle  $\theta_n$  in (4.72) and (4.73) will be randomly chosen as one of the four angles  $\{0, \pi/2, \pi, 3\pi/2\}$ . What this means is that the positions follow a lattice pattern, and this is illustrated in Fig. 4.26. The assumption is that at each time step, the molecule moves, with equal



**Fig. 4.26** Nearest neighbor random walk on a rectangular lattice



probability, to a nearest neighbor point on the lattice. Note that this is effectively the two-dimensional version of interstitial diffusion shown in Fig. 4.9.

In a similar manner as in the one-dimensional case, we let  $w(m, p, n)$  be the probability the molecule is at  $(x, y) = (m\Delta x, p\Delta y)$  after  $n$  time steps. Given that we are considering walks with a constant step size, then  $h = \Delta x = \Delta y$ . Suppose that at time step  $n$  the molecule is located at lattice point  $(m, p)$  (take, for example, the solid dot in Fig. 4.26). The molecule's position at the previous time step has to be one of the four lattice points  $(m - 1, p)$ ,  $(m + 1, p)$ ,  $(m, p - 1)$ , or  $(m, p + 1)$ , which are the hollow dots in Fig. 4.26. The probability of moving from each of these four points to  $(m, p)$  is  $\frac{1}{4}$ . We therefore have the master equation

$$w(m, p, n) = \frac{1}{4}w(m - 1, p, n - 1) + \frac{1}{4}w(m + 1, p, n - 1) + \frac{1}{4}w(m, p - 1, n - 1) + \frac{1}{4}w(m, p + 1, n - 1). \quad (4.75)$$

This is the two-dimensional equivalent of the one-dimensional master equation (4.9). The steps from this point on will follow the one-dimensional analysis very closely. To switch from  $(m, p, n)$  to  $(x, y, t)$  recall that  $x = mh$ ,  $y = ph$ , and  $t = n\Delta t$ . Introducing the function  $u(x, y, t) = w(m, p, n)$ , (4.75) takes the form

$$4u(x, t) = u(x - h, y, t - \Delta t) + u(x + h, y, t - \Delta t) + u(x, y + h, t - \Delta t) + u(x, y - h, t - \Delta t). \quad (4.76)$$

To continue we need the multivariable version of Taylor's theorem (see Appendix A). Through the quadratic terms the result is

$$\begin{aligned} f(x + h, y + \ell, t + k) &= f + \left( h \frac{\partial}{\partial x} + \ell \frac{\partial}{\partial y} + k \frac{\partial}{\partial t} \right) f + \frac{1}{2} \left( h \frac{\partial}{\partial x} + \ell \frac{\partial}{\partial y} + k \frac{\partial}{\partial t} \right)^2 f + \cdots \\ &= f + hf_x + \ell f_y + kf_t \\ &\quad + \frac{1}{2}h^2 f_{xx} + \frac{1}{2}\ell^2 f_{yy} + \frac{1}{2}k^2 f_{tt} + h\ell f_{xy} + hk f_{xt} + \ell k f_{yt} + \cdots, \end{aligned} \quad (4.77)$$

where  $f$  and its derivatives on the right-hand side are evaluated at  $(x, y, t)$ . Applying this to the terms in (4.76), and then simplifying, we obtain the following result:

$$4u = 4u - 4(\Delta t)u_t + h^2u_{xx} + h^2u_{yy} + (\Delta t)^2u_{tt} + \dots$$

Rearranging things a bit we obtain

$$u_t = \frac{h^2}{4\Delta t}(u_{xx} + u_{yy}) + \frac{\Delta t}{4}u_{tt} + \dots \quad (4.78)$$

With this, the first-order approximation for the probability satisfies

$$u_t = D(u_{xx} + u_{yy}), \quad (4.79)$$

where

$$D = \frac{h^2}{4\Delta t}. \quad (4.80)$$

The conclusion is that the resulting continuous problem is a diffusion equation. Although it was derived assuming the motion was on a rectangular lattice, as shown in Exercise 4.26, the same equation is obtained for the general random walk given in (4.72) and (4.73).

The above formula for  $D$  is a factor of two smaller than what we found for one-dimensional motion. This is not surprising if one remembers that  $D$  is a measure of the spread in the molecules per time step, and the larger  $D$  the greater the spread. In one-dimensional motion the cloud can only move left or right. In contrast, in two dimensions the particles can also move around the origin, as well as radially away from it, and this means the spreading is not as pronounced as in one dimension. In other words, the associated diffusion coefficient is less in two dimensions, and this is borne out in (4.80). Based on this, it should not be surprising that in  $d$  dimensions one still obtains a diffusion equation, with  $D = h^2/(2d\Delta t)$ .

*Example (Point Source Solution)* The symmetric solution seen in Fig. 4.25 can best be described using polar coordinates. In switching from Cartesian to polar coordinates one finds that

$$\begin{aligned} \frac{\partial}{\partial x} &= \cos\theta \frac{\partial}{\partial r} - \frac{\sin\theta}{r} \frac{\partial}{\partial \theta}, \\ \frac{\partial}{\partial y} &= \sin\theta \frac{\partial}{\partial r} + \frac{\cos\theta}{r} \frac{\partial}{\partial \theta}. \end{aligned}$$

Substituting these into (4.79), and simplifying, the polar coordinate form of the diffusion equation is

$$\frac{\partial u}{\partial t} = D \left( \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial u}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 u}{\partial \theta^2} \right). \quad (4.81)$$

We are interested in solutions that are symmetric about the origin, which is the case of the distributions shown in Fig. 4.25. Mathematically, this means that  $u$  is independent of  $\theta$ , and (4.81) reduces to

$$\frac{\partial u}{\partial t} = D \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial u}{\partial r} \right). \quad (4.82)$$

This is known as the radially symmetric diffusion equation. The second assumption, based on Fig. 4.25, is that the total number of molecules remains constant. This means that

$$\int_0^\infty \int_0^{2\pi} u r d\theta dr = \gamma.$$

Because  $u$  is independent of  $\theta$ , this reduces to

$$\int_0^\infty u r dr = \frac{\gamma}{2\pi}. \quad (4.83)$$

We want to find the solution of (4.82), that satisfies (4.83) and which also satisfies  $u \rightarrow 0$  as  $r \rightarrow \infty$ . One of the easier ways to do this is to use similarity variables. Aside from  $u$ , the only dimensional variables or parameters in the problem are  $r$ ,  $t$ ,  $D$ , and  $\gamma$ . In other words,  $u = u(r, t, D, \gamma)$ . To reduce this using dimensional analysis note  $[\gamma] = [u]/L^2$ . Using an argument very similar to the one given in Sect. 1.4, it is found that

$$u = \frac{\gamma}{Dt} F(\eta), \quad (4.84)$$

where  $\eta = r/\sqrt{Dt}$ . Substituting this into (4.82), and rearranging things a bit, yields

$$\eta F'' + F' + \frac{1}{2} \eta^2 F' + \eta F = 0.$$

This can be rewritten as

$$\frac{d}{d\eta}(\eta F') + \frac{d}{d\eta} \left( \frac{1}{2} \eta^2 F' \right) = 0.$$

Integrating, and then solving the resulting first-order differential equation for  $F$ , yields the general solution

$$F = e^{-\eta^2/4} \left( b + a \int \frac{1}{\eta} e^{\eta^2/4} d\eta \right),$$

where  $b$  and  $a$  are arbitrary constants. Now, the solution must be bounded at  $\eta = 0$ , and for this reason  $a = 0$ . The value of  $b$  is determined from (4.83), and one finds that  $b = \gamma/(4\pi)$ . Therefore, from (4.84), the solution is

$$u(x, t) = \frac{\gamma}{4\pi Dt} e^{-r^2/(4Dt)}. \quad (4.85)$$

This is an example of a point source solution with strength  $\gamma$ . ■

## 4.8 Langevin Equation

Random walks can be described as positional models of Brownian motion in the sense that they identify the locations of the molecules but they do not identify the physical reasons for the movement. To explore how to incorporate more of the physics into the modeling we need to consider what is happening to the molecule. As in Brown's original observations, the molecules involved are on the order of microns and are moving through a fluid. As such these molecules are in a sea of smaller objects, which are the atoms forming the fluid, that are undergoing thermal motions. The consequence of this is that the molecule is constantly subjected to many random impacts from these rapidly moving smaller objects. Although each atom has only a small effect on the molecule, there are many of them and together they are responsible for the molecule's random motion.

To model this we will use Newton's second law, namely  $\mathbf{F} = m\mathbf{a}$ . The force  $\mathbf{F}$  between the molecule and the surrounding fluid will be separated into a deterministic component  $\mathbf{D}$ , and a random component  $\mathbf{R}$ , and we write

$$\mathbf{F} = \mathbf{D} + \mathbf{R}. \quad (4.86)$$

Determining  $\mathbf{D}$  and  $\mathbf{R}$  is based on the observation that the relevant time and space scales for the molecule and surrounding atoms are very different. The thermal motions of the atoms are rapid, and occur over very short distances, compared to those for the molecule. The force  $\mathbf{F}$  accounts for the multiple individual collisions that are taking place as the molecule moves. The first term,  $\mathbf{D}$ , is the resistance force. As the molecule moves through the fluid there will be more collisions with the surrounding atoms on the front than on the back, and this will give rise to a resistance force. This is analogous to air resistance experienced by an object falling in air, and it is accounted for in (4.86) with  $\mathbf{D}$ . It is assumed that this force is proportional to the velocity. Letting  $\mathbf{r}(t)$  be the position of the molecule, then  $\mathbf{D} = -\mu\mathbf{r}'$  where  $\mu$  is a constant. The term  $\mathbf{R}$  is suppose to account for everything else the atoms are doing to the molecule. As such it contains the random, and rapidly fluctuating, component of the force. The resulting equation of motion is

$$m \frac{d^2 \mathbf{r}}{dt^2} = -\mu \frac{d\mathbf{r}}{dt} + \mathbf{R}(t), \quad (4.87)$$

where  $m$  is the mass of the molecule. This is known as the *Langevin equation*. It is an example of a stochastic differential equation due to the presence of  $\mathbf{R}$ . As a mathematical model it has been very influential in classical and quantum mechanics, as well as in statistical mechanics. In fact, Langevin ideas remain fundamental to contemporary scientific research in nonequilibrium statistical physics.

To continue it is necessary to specify  $\mathbf{R}$ . Although it does fluctuate rapidly, the amplitude or magnitude of this function is not small. In fact, to quote Langevin, “it maintains the agitation” of the molecule (Lemons and Gythiel, 1997). What he means is that  $\mathbf{R}$  is the driving force that is responsible for the observed random walk behavior of the molecule. This is interesting information, but it still leaves open the question of how to determine  $\mathbf{R}$ . The usual approach is to employ results from probability theory to write down the defining formulas for  $\mathbf{R}$ . The approach used here is more fundamental, and the formulas are derived directly from the properties of Brownian motion.

One of the more basic hypotheses is that  $\mathbf{R}$  is an external force that is independent of the molecule’s motion, in other words, it does not depend on  $\mathbf{r}$  or its derivatives. This assumption enables us to solve the equation. Introducing the velocity  $\mathbf{v} = \mathbf{r}'$ , then (4.87) can be written as

$$\frac{d\mathbf{v}}{dt} + \lambda\mathbf{v} = \frac{1}{m}\mathbf{R}(t), \quad (4.88)$$

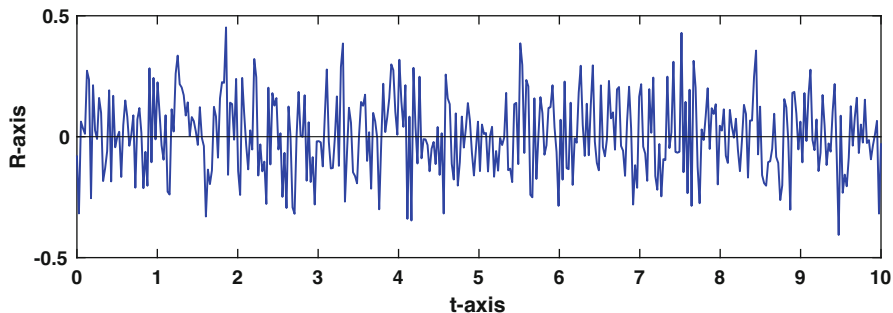
where  $\lambda = \mu/m$ . This first-order equation can be solved using an integrating factor, and the result is

$$\mathbf{v}(t) = \mathbf{v}(0)e^{-\lambda t} + \frac{1}{m} \int_0^t \mathbf{R}(\tau)e^{-\lambda(t-\tau)} d\tau. \quad (4.89)$$

This shows that the random forcing has a cumulative effect on the velocity because it depends on an integral of  $\mathbf{R}$ . How much the early values of  $\mathbf{R}$  affect  $\mathbf{v}$  depends on  $\lambda$ . The larger  $\lambda$ , the greater the exponential decay in the integral and the less effect the early values of  $\mathbf{R}$  have on the velocity. Also note that larger values of  $\lambda$  reduce the contribution of the initial velocity. Said another way, the larger  $\lambda$  is, the quicker the molecule forgets its initial velocity and its movement is determined by Brownian randomization. Once the velocity is known, the position of the molecule can be determined by integrating (4.89), and the result is

$$\mathbf{r}(t) = \mathbf{r}(0) + \frac{1}{\lambda}\mathbf{v}(0)(1 - e^{-\lambda t}) + \frac{1}{m\lambda} \int_0^t \mathbf{R}(\tau)(1 - e^{-\lambda(t-\tau)}) d\tau. \quad (4.90)$$

Stating that the forcing function  $\mathbf{R}$  is random does not mean that it is arbitrary. To be consistent with Brownian motion,  $\mathbf{R}$  is subject to certain restrictions, and these will be derived in the next section. Before that, a comment is needed about the mathematical problem we are addressing. The example of the random forcing term  $\mathbf{R}$  shown in Fig. 4.27 uses 400 points along the  $t$ -axis. As will be explained later, the



**Fig. 4.27** Example of the random, rapidly fluctuating, function  $\mathbf{R}$  appearing in the Langevin equation (4.87)

value of  $\mathbf{R}(t_1)$  is assumed to be independent of the value of  $\mathbf{R}(t_2)$  if  $t_1 \neq t_2$ . This means that if more than 400 points are used, the graph will appear even more random than in Fig. 4.27. The question that immediately arises is whether the resulting nondifferentiability of this function causes the differential equation (4.88), or its solution (4.88), to be meaningless. The answer as to why it is possible to include such a forcing function is one of the central objectives of stochastic differential equations, and how this is done is explained in Appendix C. The short answer is that differentiability is not an issue in (4.89) or (4.90), and it is these expressions that we will work with.

### 4.8.1 Properties of the Random Forcing

The solution in (4.89) is for a single molecule. We are interested in what happens when a large group of molecules are released at a point, which we will assume is the origin. Also, for simplicity, it is assumed that the molecules start out at rest, so  $\mathbf{v}(0) = \mathbf{0}$ . If there are  $K$  molecules in the group, then the mean velocity of the group is

$$\mathbf{V} = \frac{1}{K} \sum_{i=1}^K \mathbf{v}_i,$$

where  $\mathbf{v}_i$  is the velocity of the  $i$ th molecule. Similarly, the mean displacement of the group is

$$\mathbf{M} = \frac{1}{K} \sum_{i=1}^K \mathbf{r}_i,$$

where  $\mathbf{r}_i$  is the displacement of the  $i$ th molecule. Using (4.89) we have that

$$\mathbf{V} = \frac{1}{m} \int_0^t \mathbf{Q}(\tau) e^{-\lambda(t-\tau)} d\tau,$$

and from (4.90)

$$\mathbf{M} = \frac{1}{m\lambda} \int_0^t \mathbf{Q}(\tau) (1 - e^{-\lambda(t-\tau)}) d\tau, \quad (4.91)$$

where

$$\mathbf{Q} = \frac{1}{K} \sum_{i=1}^K \mathbf{R}_i \quad (4.92)$$

is the mean random force, and  $\mathbf{R}_i$  is the random forcing for the  $i$ th molecule.

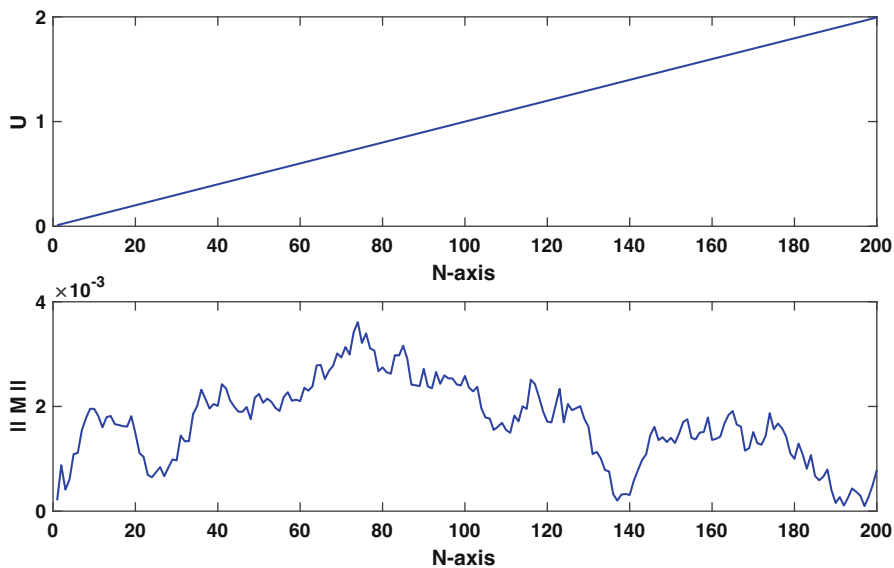
*Assumption 1: Zero Average*

As you might have already noticed, the molecules in the group are identical, so they have the same mass  $m$  and resistance factor  $\mu$ . This brings us to the first assumption made on the random forcing. It is perhaps easiest to explain this using the two-dimensional random walk in (4.72) and (4.73). All directions are equally likely. So, at time step  $n$ , if we happen to select a direction angle  $\theta_n$  we could have just as likely selected the opposite direction, either  $\theta_n + \pi$  or  $\theta_n - \pi$ . Consequently, at any given time step  $n$ , if one averages over all the displacements possible they get zero. The random forcing is assumed to be consistent with this result. In other words, it is assumed that when letting  $K \rightarrow \infty$  in (4.92) they obtain  $\mathbf{Q} = \mathbf{0}$ . With this, taking the same limit in (4.91), we have that  $\mathbf{M} = \mathbf{0}$ . This result does not mean the group is motionless, rather it means the motion is symmetric. Given that it is equally likely to move in one direction as another, when a very large group starts out at the same point, then the group will be distributed approximately symmetrically about the point as time progresses. Consequently, the resulting average displacement is approximately zero. This symmetry is evident in the scatter plots in Fig. 4.2, as well as in the distributions in Fig. 4.25.

*Assumption 2: Independence*

It is known that for random walks in one dimension the average displacement is zero, but the average of the displacement squared grows linearly in time (see Exercise 4.1). This same conclusion holds for multidimensional random walks, and to investigate this for the Langevin equation let

$$U = \frac{1}{K} \sum_{i=1}^K \mathbf{r}_i \cdot \mathbf{r}_i. \quad (4.93)$$



**Fig. 4.28** The upper graph gives the distance squared (4.93) averaged over a group of molecules moving according to the random walk (4.72) and (4.73). The lower graph gives  $||\mathbf{M}||$ . In the calculation,  $K = 100,000$  molecules were used, and the step length was 0.1

The values of  $U$  are given in Fig. 4.28 for the two-dimensional random walk (4.72) and (4.73). For completeness, the values of the magnitude of the average displacement vector (4.91) are also given. In looking at the values for  $||\mathbf{M}||$  one might be a bit skeptical about the statement that the average displacement is zero. It should be remembered that this holds in the limit of  $K \rightarrow \infty$ . Also, after 200 time steps the molecules have the potential to be a distance of  $200h = 20$  from the origin. Compared to this, the values for  $||\mathbf{M}||$  in Fig. 4.28 are quite small. No such qualifications, however, need to be made about the computed values of  $U$ , which clearly show a linear dependence on time.

The question we now consider is whether the Langevin equation results in  $U$  increasing linearly in time. Substituting the solution (4.90) into (4.93), and recalling that  $\mathbf{r}_i(0) = \mathbf{v}_i(0) = \mathbf{0}$ , we obtain

$$U = \frac{1}{K} \sum_{i=1}^K \int_0^t \int_0^t \mathbf{R}_i(s) \cdot \mathbf{R}_i(\tau) f(s, \tau) d\tau ds, \quad (4.94)$$

where

$$f(s, \tau) = \frac{1}{m^2 \lambda^2} (1 - e^{-\lambda(t-\tau)})(1 - e^{-\lambda(t-s)}). \quad (4.95)$$



This brings us to the next assumption made on the random forcing. Compared to the molecule's motion, the surrounding atoms are moving very quickly, and they are undergoing a large number of collisions with their neighbors over a very short amount of time. Consequently, the atomic events responsible for the random force at time  $t$  are effectively independent of those for the random force at a different time  $\tau$ . In this case the forcing function is said to be Markovian. A consequence of this assumption is that the positive and negative values of  $\mathbf{R}_i(s) \cdot \mathbf{R}_i(\tau)$  are all equally likely. It is for this reason that the average of  $\mathbf{R}_i(s) \cdot \mathbf{R}_i(\tau)$ , as  $K \rightarrow \infty$ , is zero if  $s \neq \tau$ . However, this does not mean that  $U \rightarrow 0$  as  $K \rightarrow \infty$  in (4.94) because we need to consider what happens when  $s = \tau$ .

*Assumption 3: Concentration*

Assuming that the forcing is nonzero, the product  $\mathbf{R}_i(s) \cdot \mathbf{R}_i(s)$  is positive. This means that the random forcing tends to accentuate the values in the integrals in (4.94) for  $s = \tau$ . The specific assumption made is that given any continuous function  $f(s, \tau)$ , if  $0 < s < t$ , then

$$\lim_{K \rightarrow \infty} \frac{1}{K} \sum_{i=1}^K \int_0^t \mathbf{R}_i(s) \cdot \mathbf{R}_i(\tau) f(s, \tau) d\tau = \gamma f(s, s), \quad (4.96)$$

where  $\gamma$  is a positive constant. As the above equation shows,  $\sqrt{\gamma}$  is the amplitude of the forcing, and its value will be determined below when comparing the random walk and Langevin descriptions.

With (4.96), letting  $K \rightarrow \infty$  in (4.94) we have that

$$\begin{aligned} U &= \frac{1}{m^2 \lambda^2} \int_0^t \gamma \left(1 - e^{-\lambda(t-s)}\right)^2 ds \\ &= \frac{\gamma}{2m^2 \lambda^3} \left(2\lambda t - 3 + 4e^{-\lambda t} - e^{-2\lambda t}\right). \end{aligned} \quad (4.97)$$

For large values of time the above solution reduces to

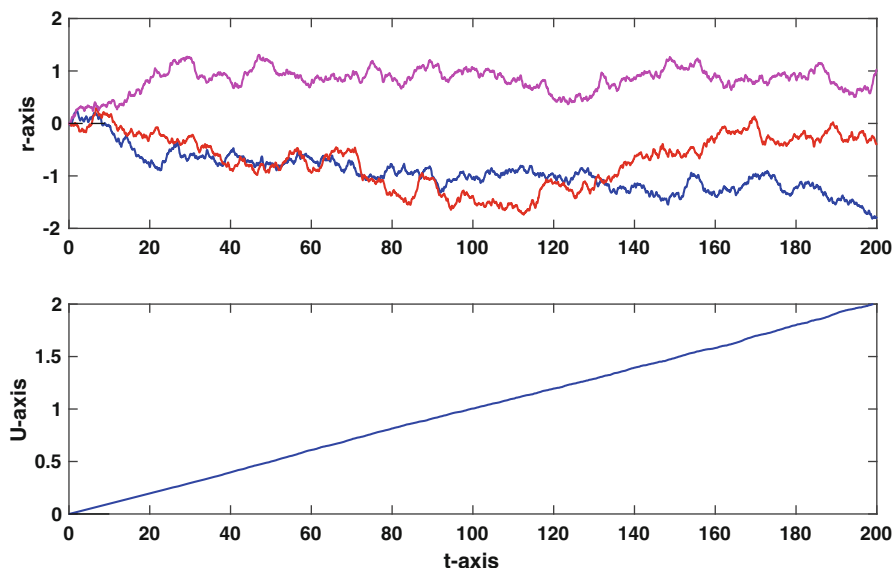
$$U \approx \frac{\gamma}{\mu^2} t, \quad (4.98)$$

where  $\mu$  is the damping coefficient in (4.87). Therefore, with the stated assumptions on the random forcing, the Langevin equation gives us the expected conclusion that  $U$  increases linearly in time.

The two constants in (4.98) can be determined by introducing additional assumptions into the formulation. It is commonly assumed that the resistance term  $\mu \mathbf{r}'$  in the Langevin equation is equivalent to the viscous force of a fluid. As shown in (1.22), it is known that for slow flows the drag force on a sphere of radius  $r$  is  $6\pi \nu r v$ , where  $\nu$  is the dynamic viscosity of the fluid and  $v$  is the velocity of the sphere. This is Stokes' law for the drag on a sphere and was introduced in Sect. 1.2.2. Assuming that the molecules are spheres, then the conclusion is that

$\mu = 6\pi\eta$ . Because the viscosity of fluids such as air and water is known, then the corresponding value of  $\mu$  is known. The value of  $\gamma$  can be determined from the theory of the kinetic theory of gases because the integral in (4.97) is associated with the thermal energy of the system. It is found that  $\gamma = 6\mu kT$ , where  $k$  is the Boltzmann constant and  $T$  is the absolute temperature. To relate this to the diffusion process arising from the random motion, it is known that  $U = 2dDt$ , where  $d$  equals the number of spatial dimensions (see Exercise 4.1). Using (4.98) we have that  $D = \gamma/(2d\mu^2)$ . Combining this with our values for the two constants we obtain the Stokes-Einstein equation (4.15).

*Example* Because of the random forcing, each time the Langevin equation is solved a different solution is obtained. Three such solutions  $r(t)$  are shown in the upper graph of Fig. 4.29 for one-dimensional motion. For each solution the initial conditions are  $r(0) = 0$  and  $r'(0) = 0$ , and the parameters are  $\mu = 10$ ,  $\gamma = 1$ , and  $m = 1$ . The curves show the typical wandering of a Brownian motion. To check that the distance squared (4.93) is linear, the values of  $U(t)$  are plotted in the lower graph of Fig. 4.29. Not only is the curve linear, it is the same curve obtained using random walks shown in Fig. 4.28 assuming  $\Delta t = 1$ . This is not a coincidence. From the Langevin formulation we have that  $D = \gamma/(2d\mu^2)$ , while from the random walk we have  $D = h^2/(2d\Delta t)$ . For these to produce the same diffusion coefficient it is therefore required that  $\gamma = (\mu h)^2/\Delta t$ . The given values of  $\gamma$ ,  $\mu$ , and  $h$  satisfy this equation, and that is why Fig. 4.29 agrees with Fig. 4.28. ■



**Fig. 4.29** The upper graph gives three solutions of the one-dimensional Langevin equation, assuming  $r(0) = 0$  and  $r'(0) = 0$ . The lower graph gives the distance squared (4.93), averaged over 10,000 solutions of the Langevin equation

*Example (Asset Modeling)* The Langevin equation was derived using ideas from molecular physics, but it has application in a wide variety of areas. One is in modeling the value of a financial asset, such as a stock. To frame this in terms of the discrete time steps used for a random walk, suppose the value of the asset at  $t = n\Delta t$  is  $V_n$  and we want to determine its value  $V_{n+1}$  at the next time step  $t = (n+1)\Delta t$ . The assumption is that the asset changes by an amount proportional to its value, and so

$$V_{n+1} = V_n + r_n V_n. \quad (4.99)$$

The coefficient  $r_n$  is the rate of return. For example, if the asset is a simple savings account, and the interest rate is  $\mu$ , then  $r_n = \mu\Delta t$ . The value of many assets, such as stocks, are affected by external events, and their rates can vary dramatically with time. To account for this in the model, the rate is assumed to have the form

$$r_n = \mu\Delta t + \sigma\Delta W, \quad (4.100)$$

where the terms in this expression are explained below.

- *Expected Average Growth.* If external events do not affect the asset, then its value is assumed to increase at a constant rate. Just as with the savings account example, this rate is assumed to be  $\mu\Delta t$ . The positive constant  $\mu$  is known as the drift coefficient.
- *Random Fluctuations.* The value of a stock can change due to rapidly changing external events. The  $\sigma\Delta W$  term in (4.100) accounts for these fluctuations. In this expression,  $\sigma$  is a positive constant that depends on the particular asset under study, and is known as the volatility. The random function  $\Delta W$  is time-dependent, but independent of the asset.

Combining (4.99) and (4.100) we have that

$$V_{n+1} - V_n = \mu\Delta t V_n + \sigma\Delta W V_n, \quad (4.101)$$

or equivalently

$$\frac{V_{n+1} - V_n}{\Delta t} = \left( \mu + \sigma \frac{\Delta W}{\Delta t} \right) V_n. \quad (4.102)$$

It is tempting to let  $\Delta t \rightarrow 0$  in this expression, and from this conclude that  $V' = (\mu + \sigma R)V$ , where  $R = W'$ . However, the existence of this limit for the product  $\Delta W V_n$  is questionable. One way to avoid this is to change variables and transform it into a Langevin equation. With this in mind, assume the change of variables has the form  $Q = f(V)$ . In this case, using Taylor's theorem for small  $\Delta t$ ,

$$\begin{aligned} Q_{n+1} &= f(V(t_n + \Delta t)) \\ &= f\left(V_n + \Delta t V'_n + \frac{1}{2}(\Delta t)^2 V''_n + \dots\right) \end{aligned}$$

$$\begin{aligned}
&= f_n + \left( \Delta t V'_n + \frac{1}{2}(\Delta t)^2 V''_n + \cdots \right) f'_n + \frac{1}{2}(\Delta t)^2 (V'_n)^2 f''_n + \cdots \\
&= Q_n + \Delta t (\mu + \sigma R_n) V_n f'_n + \frac{1}{2}(\Delta t)^2 (\mu + \sigma R_n)^2 V_n^2 f''_n + \cdots \quad (4.103)
\end{aligned}$$

For the random forcing associated with Brownian motion, it can be shown that for small  $\Delta t$ ,  $R^2 = 1/\sqrt{\Delta t} + \cdots$ . Also, to transform (4.101) into one that resembles the Langevin equation let  $Q = \ln(V)$ . With this, (4.103) becomes

$$Q_{n+1} - Q_n = \Delta t \left( \mu - \frac{1}{2}\sigma^2 \right) + \sigma \Delta t R_n. \quad (4.104)$$

Letting  $\Delta t \rightarrow 0$  we obtain

$$\frac{dQ}{dt} = \mu - \frac{1}{2}\sigma^2 + \sigma R, \quad (4.105)$$

Letting  $W = \int_0^t R(\tau) d\tau$ , then the solution is

$$Q(t) = Q(0) + \left( \mu - \frac{1}{2}\sigma^2 \right) t + \sigma W(t).$$

Transforming back into the original variables,

$$V(t) = V(0)e^{\lambda t + \sigma W(t)}, \quad (4.106)$$

where  $\lambda = \mu - \frac{1}{2}\sigma^2$ . This solution is an example of what is known as geometric Brownian motion. ■

### 4.8.2 Endnotes

The assumption that  $\mathbf{R}$  in (4.87) is a randomly varying function is an approximation. The reasoning used when introducing randomness is that this reflects the zigzag nature of the motion. For the space and time scales we were considering, this is appropriate. However, if you were to slow time down, and look at the molecular level, the motion would appear to be smooth. As an example, if you watch a slowed down movie of billiard balls bouncing off each other, at the level of the billiard balls, the motion would appear smooth. Yet, in real-time they give the impression of changing directions instantly on impact. What this means is that the non-differentiability of the function in Fig. 4.27 is a consequence of the approximation of the forcing function. This observation has had a significant impact on the development of stochastic differential equations, and it specifically relates to how

the integrals in (4.89) and (4.90) are defined. In one formulation the integrals possess important mathematical properties expected of integrals, but are not completely consistent with the physics, while other formulations do just the opposite. Exploring the ramifications of this statement is beyond the scope of this text, and those who want to learn more about this should consult Mazo (2002) and Kampen (2007).

## Exercises

### Sections 4.2 and 4.3

**4.1** Suppose a total of  $K$  particles, all starting at  $x = 0$ , undergo a random walk. Let  $x_i(n)$  be the position of the  $i$ th particle at time step  $n$ .

- (a) Writing  $x_i(n) = x_i(n-1) + q_i(n)$ , explain why the value of  $q_i$  is either  $\Delta x$  or  $-\Delta x$ .
- (b) Use the basic properties of a random walk to explain why the following holds:

$$\lim_{K \rightarrow \infty} \frac{1}{K} \sum_{i=1}^K q_i(n) = 0.$$

Explain why the same reasoning can be used to explain why, if  $n \neq k$ , then

$$\lim_{K \rightarrow \infty} \frac{1}{K} \sum_{i=1}^K q_i(n)q_i(k) = 0.$$

What is the above limit in the case of when  $k = n$ ?

- (c) The mean displacement of the group, at time step  $n$ , is

$$d_K(n) = \frac{1}{K} \sum_{i=1}^K x_i(n).$$

Relate  $d_K(n)$  with  $d_K(n-1)$ , and from this show that  $\lim_{K \rightarrow \infty} d_K = 0$ . Therefore, the average displacement of a large group of particles is approximately zero.

- (d) The mean-square displacement of the group is defined as

$$r_K(n) = \frac{1}{K} \sum_{i=1}^K x_i^2(n).$$

The value of  $r_K$  is a measure of the spread of the group. By relating  $x_i(n)$  with  $x_i(n-1)$ , show that  $\lim_{K \rightarrow \infty} r_K = n(\Delta x)^2$ . Therefore, on average, for very

large groups of particles the mean-square displacement of the group increases linearly with time.

- (e) There are various ways to derive the formula for the diffusion coefficient. One sometimes used in physics is

$$D = \frac{\langle x^2 \rangle}{2t},$$

where  $\langle x^2 \rangle = \lim_{K \rightarrow \infty} r_K$ . Show why this agrees with the definition in (4.12).

**4.2** This problem makes use of the connection of the diffusion coefficient with the molecule's mean free path and average time between collisions, as expressed by the Einstein-Smoluchowski equation (4.14). In what follows the average speed of the molecule is defined as  $v = \lambda/\tau$ .

- (a) In air the mean free path is in the neighborhood of 30 times the average molecular separation distance  $d$ . In air suppose  $d$  is approximately  $3 \times 10^{-7}$  cm. Given that for air  $D \approx 0.2 \frac{\text{cm}^2}{\text{s}}$ , approximately how long is it between collisions? Approximately how fast are the molecules traveling? How many collisions are there per second?
- (b) In water the mean free path is in the neighborhood of 30 times the average molecular separation distance  $d$ . In water suppose  $d$  is approximately  $3 \times 10^{-8}$  cm. Given that for water  $D \approx 2 \times 10^{-5} \frac{\text{cm}^2}{\text{s}}$ , approximately how long is it between collisions? Approximately how fast are the molecules traveling? How many collisions are there per second?

**4.3** In a random walk suppose that at  $t = (n - 1)\Delta t$  the particle is located at  $x = m\Delta x$ . The assumption is that at  $t = n\Delta t$  the particle will have moved to  $x = (m + 1)\Delta x$  with probability  $3/4$  or to  $x = (m - 1)\Delta x$  with probability  $1/4$ .

- (a) Draw a grid that shows the achievable positions a particle can reach at  $n = 0, 1, 2, 3$  when starting at  $(m, n) = (0, 0)$ . Letting  $w(m, n)$  be the probability the particle is at  $x = m\Delta x$  after  $n$  time steps, determine  $w(m, n)$  for the positions shown in the grid.
- (b) Based on your result in part (a) what are the values of  $A, B$  so  $w(m, n) = Aw(m - 1, n - 1) + Bw(m + 1, n - 1)$ .
- (c) Setting  $u(x, t) = w(m, n)$ , rewrite your result from part (b) in terms of  $u(x, t)$ . Assuming  $\Delta x$  and  $\Delta t$  are small, derive a partial differential equation for  $u$ .

**4.4** In a random walk suppose that at  $t = (n - 1)\Delta t$  the particle is located at  $x = m\Delta x$ . The assumption is that at  $t = n\Delta t$  the particle will have moved to  $x = (m + 2)\Delta x$  with probability  $1/3$  or to  $x = (m - 1)\Delta x$  with probability  $2/3$ .

- (a) Draw a grid that shows the achievable positions a particle can reach at  $n = 0, 1, 2, 3$  when starting at  $(m, n) = (0, 0)$ . Letting  $w(m, n)$  be the probability the particle is at  $x = m\Delta x$  after  $n$  time steps, determine  $w(m, n)$  for the positions shown in the grid.

- (b) Based on your result in part (a) what are the values of  $A, B$  so  $w(m, n) = Aw(m-2, n-1) + Bw(m+1, n-1)$ .
- (c) Setting  $u(x, t) = w(m, n)$ , rewrite your result from part (b) in terms of  $u(x, t)$ . Assuming  $\Delta x$  and  $\Delta t$  are small, derive a partial differential equation for  $u$ .

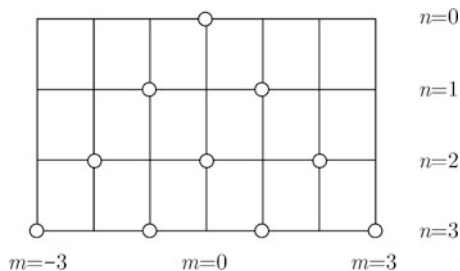
**4.5** A lazy random walk is one that allows the particle to stay put instead of having to move left or right. For this situation assume the probability of going to the right is  $p_r$ , the probability of going to the left is  $p_\ell$ , and the probability of not moving is  $p_s$ . As usual,  $p_\ell + p_s + p_r = 1$ . Also, letting  $\Delta x$  be the spatial stepsize and  $\Delta t$  the temporal stepsize, then  $x = m\Delta x$  and  $t = n\Delta t$ .

- (a) Draw a grid that shows the achievable positions a particle can reach at  $n = 0, 1, 2$  when starting at  $(m, n) = (0, 0)$ . Letting  $w(m, n)$  be the probability the molecule is at  $x = m\Delta x$  after  $n$  time steps, then we know that  $w(-1, 1) = p_\ell w(0, 0) = p_\ell$ ,  $w(0, 1) = p_s w(0, 0) = p_s$ , and  $w(1, 1) = p_r w(0, 0) = p_r$ . Determine the probability for the positions at  $n = 2$ . Also, show that the probabilities at each time level ( $n = 0, 1, 2$ ) add to one and explain why this has to be the case.
- (b) Based on your result in part (a), what are the values of  $A, B$ , and  $C$ , so  $w(m, n) = Aw(m-1, n-1) + Bw(m, n-1) + Cw(m+1, n-1)$ .
- (c) Setting  $u(x, t) = w(m, n)$ , then rewrite your result from part (b) in terms of  $u$ . Assuming  $\Delta x$  and  $\Delta t$  are small, and assuming that  $p_\ell = p_r$ , derive a partial differential equation for  $u$ . The coefficient in the equation must not depend on  $x$  or  $t$ , but it can depend on  $\Delta x, \Delta t, p_\ell, p_s, p_r$ , etc.
- (d) Explain how this result differs from (4.12) and (4.13). If one diffusion coefficient is smaller than the other, provide a physical explanation why this happens.

**4.6** For the random walk we considered there was no memory of the previous step when determining the current one. An interesting modification is a correlated walk where the probability at the next time step depends upon the previous step. To examine this suppose that step  $n-1$  is complete. Step  $n$  is made in the same direction with probability  $p$  and in the opposite direction with probability  $1-p$ . If  $p > \frac{1}{2}$  it is called a persistent walk and if  $p < \frac{1}{2}$  it is an anti-persistent walk. To get this procedure started, use a regular random walk at the first step. Also, assume that  $0 < p < 1$ .

- (a) The grid in Fig. 4.30 identifies the achievable positions a particle can reach at each time step when starting at  $(m, n) = (0, 0)$ . Letting  $w(m, n)$  be the probability the molecule is at  $x = m\Delta x$  after  $n$  time steps, then  $w(-1, 1) = 1/2$ , and  $w(1, 1) = 1/2$ . Determine the probability for the other positions shown in the figure.
- (b) Let  $w(m, n) = f(m, n) + g(m, n)$ , where  $f(m, n)$  is the probability of arriving at  $x = m\Delta x$ , at time step  $n$ , from the left, and  $g(m, n)$  is the probability of arriving at  $x = m\Delta x$ , at time step  $n$ , from the right. Show that  $f(m, n) = pf(m-1, n-1) + (1-p)g(m-1, n-1)$  and  $g(m, n) = (1-p)f(m+1, n-1) + pg(m+1, n-1)$ .

**Fig. 4.30** Figure for Exercises 4.6, 4.7, and 4.8



- (c) Setting  $u(x, t) = f(m, n) + g(m, n)$  and  $v(x, t) = f(m, n) - g(m, n)$ , expand  $u, v$  for small  $\Delta x, \Delta t$ . Letting  $\Delta t \rightarrow 0$ , with  $c = \Delta x / \Delta t$  fixed and  $p = 1 - \alpha \Delta t / 2$ , derive the following partial differential equation

$$u_{tt} + \alpha u_t = c^2 u_{xx}.$$

This is known as the telegraph equation.

- (d) As in part (b), expand  $u, v$  for small  $\Delta x, \Delta t$ , but now let  $\Delta t \rightarrow 0$  with  $\Delta x^2 / \Delta t$  fixed and  $p$  constant. Show that  $u$  satisfies the diffusion equation, where the diffusion coefficient is

$$D = \frac{\Delta x^2}{2\Delta t} \frac{p}{1-p}.$$

**4.7** A random walk with loss is one that allows the particle to be irreversibly lost from the system at each time step. Suppose that at  $t = (n-1)\Delta t$  the particle is located at  $x = m\Delta x$ . The assumption is that at  $t = n\Delta t$  the particle will have moved to  $x = (m+1)\Delta x$  with probability  $p_r$ , it will have moved to  $x = (m-1)\Delta x$  with probability  $p_r$ , and it will have vanished with probability  $p_s$ . As usual,  $2p_r + p_s = 1$ .

- (a) The grid in Fig. 4.30 identifies the achievable positions a particle can reach at each time step when starting at  $m = 0$ . Letting  $w(m, n)$  be the probability the molecule is at  $x = m\Delta x$  after  $n$  time steps, then we know that  $w(-1, 1) = p_r w(0, 0) = p_r$ , and  $w(1, 1) = p_r w(0, 0) = p_r$ . Determine the probability for the other positions shown in the figure.
- (b) Based on your result in part (a), what are the values of  $A, B$  and  $C$ , so  $w(m, n) = Aw(m-1, n-1) + Bw(m, n-1) + Cw(m+1, n-1)$ ?
- (c) Setting  $u(x, t) = w(m, n)$ , rewrite your result from part (b) in terms of  $u(x, t)$ . Assuming  $\Delta x, \Delta t$  are small, derive a partial differential equation for  $u$ . In doing this assume the probability of loss is small, that is, assume  $p_s = p_0 \Delta t$ . The coefficients of the equation you derive must not depend on  $x$  or  $t$  but can depend on  $\Delta x, \Delta t, p_r$ , and  $p_0$ . Also, in the case of when  $p_0 = 0$  your result should reduce to the diffusion equation (4.13).



**4.8** This problem considers a random walk when the probabilities of left or right steps are not equal. In particular the probability of going right is  $p_r$  and the probability of going to the left is  $p_\ell$ . It is assumed that  $p_r, p_\ell$  are positive and  $p_r + p_\ell = 1$ .

- (a) The grid in Fig. 4.30 identifies the achievable positions a particle can reach at each time step. Letting  $w(m, n)$  be the probability of each position, then we know that  $w(-1, 1) = p_\ell$ ,  $w(0, 1) = 0$  and  $w(1, 1) = p_r$ . One can show that  $w(0, 2) = p_r w(-1, 1) + p_\ell w(1, 1) = 2p_r p_\ell$ , that is,  $w(0, 2)$  is the sum of the probability of moving right from  $(-1, 1)$  and moving left from  $(1, 1)$ . Use this principle to determine the probability for the other positions shown in the figure below. Also, show that the probabilities at each time level ( $n = 0, 1, 2, 3$ ) add to one and explain why this has to be the case.
- (b) In going from time level  $n = 0$  to time level  $n \neq 0$ , explain why to reach  $x = m\Delta x$  it takes  $n_r = (n + m)/2$  steps to the right and  $n_\ell = (n - m)/2$  steps to the left.
- (c) There are  $n!/(n_r!n_\ell!)$  unique paths to reach  $x = m\Delta x$  and as a consequence of this

$$w(m, n) = (p_r)^{n_r} (p_\ell)^{n_\ell} \frac{n!}{n_r!n_\ell!},$$

for  $m = -n, -n + 2, -n + 4, \dots, n$ . Verify this formula for the positions shown in Fig. 4.30.

- (d) Use Stirling's approximation to write  $w$  so it has the form  $\alpha M^{n_\ell} N^{n_r}$ , where  $N$  is written in terms of  $p_r, n$ , and  $m$ , while  $M$  is written in terms of  $p_\ell, n$ , and  $m$ . Setting  $Q = \ln(M^{n_\ell} N^{n_r})$  find the  $m$  that maximizes  $Q$ . After this, use Taylor's theorem, through quadratic terms, to expand  $Q$  around this  $m$  value. From this show that for large  $n$ ,

$$w(m, n) \sim \frac{1}{\sqrt{2\pi n p_r p_\ell}} e^{-z},$$

where

$$z = \frac{[m - n(p_r - p_\ell)]^2}{8n p_r p_\ell}.$$

- (e) Use the principle in part (a) to derive a master equation, which expresses  $w(m, n)$  in terms of  $w(m - 1, n - 1)$  and  $w(m + 1, n - 1)$ . Setting  $u(x, t) = w(m, n)$ , then rewrite your equation in terms of  $u$ .
- (f) Expand the equation for  $u$  from part (e) for small  $\Delta x$  and  $\Delta t$ . Assuming that  $p_r = p_\ell + \lambda \Delta x$ , derive a partial differential equation for  $u$  that involves  $u_x, u_t, u_{xx}$ . The coefficients of the equation must not depend on  $x$  or  $t$  but can depend on  $\lambda, \Delta x, \Delta t, p_r, p_\ell$ , etc. The equation you are deriving is called the drift-

diffusion equation. How this equation can be derived without the assumption made about  $p_r$  can be found in Holmes (2013a).

- (g) Show that the solution in part (f) can be written, up to a multiplicative constant, as

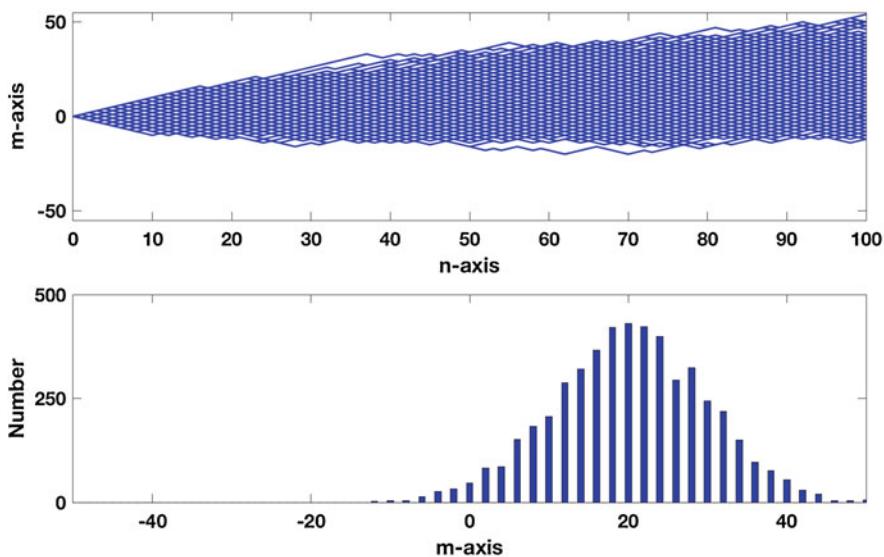
$$u(x, t) = \frac{1}{\sqrt{t}} e^{-(x-vt)^2/(at)}.$$

Show that this satisfies your drift-diffusion equation and in the process state how  $v$  and  $\alpha$  are related to  $\lambda$ ,  $\Delta x$ ,  $\Delta t$ ,  $p_r$ ,  $p_\ell$ .

- (h) Explain why the constant  $v$  in part (g) is known as the drift velocity.  
 (i) Figure 4.31 shows the result of running a biased random walk with 5000 particles all starting at the origin. From these data, and your result in part (d), estimate the values of  $p_r$  and  $p_\ell$ .

**4.9** For (4.10), when using Taylor's theorem for small  $\Delta x$ ,  $\Delta t$ , suppose you include terms up to third-order.

- (a) How does (4.11) change?  
 (b) Using the same three possibilities for the ratio  $(\Delta x)^2/\Delta t$ , show that (4.13) is obtained.



**Fig. 4.31** Figure for Exercise 4.8

## Section 4.4

**4.10** For the diffusion problem in (4.31) and (4.32) suppose

$$f(x) = \frac{\alpha}{\sqrt{\pi}} e^{-\alpha^2 x^2},$$

where  $\alpha$  is a positive constant.

(a) Show that

$$u(x, t) = \frac{1}{2\sqrt{\pi D(t + \tau)}} e^{-\frac{x^2}{4D(t + \tau)}},$$

where  $\tau = 1/(4\alpha^2 D)$ , satisfies the diffusion equation and the stated initial condition.

(b) Show that the solution in part (a) satisfies  $\int_{-\infty}^{\infty} u(x, t) dx = 1$ .

(c) Explain how the result in part (b) can be obtained directly from the diffusion equation.

**4.11** This problem considers the solution for one or more point sources.

(a) Show that

$$u_1(x, t) = P_1 \frac{1}{2\sqrt{\pi D t}} e^{-(x-x_1)^2/(4Dt)},$$

satisfies the diffusion equation. Also, show that, for  $t > 0$ ,

$$\int_{-\infty}^{\infty} u_1(x, t) dx = P_1,$$

and that  $u_1 \rightarrow 0$  for  $t \rightarrow 0^+$  if  $x \neq x_1$ . This is known as the solution of the diffusion equation with a point source of strength  $P_1$  at  $x = x_1$ .

(b) Suppose that  $u_1(x, t)$  is the point source solution, of strength  $P_1$  at  $x = x_1$ , and  $u_2(x, t)$  is the point source solution, of strength  $P_2$  at  $x = x_2$ . Use  $u_1$  and  $u_2$  to write down the solution  $u$  of the diffusion equation that has point sources at  $x_1$  and  $x_2$ , with respective strengths  $P_1$  and  $P_2$ . You should show that  $u$  satisfies the diffusion equation, that  $u \rightarrow 0$  for  $t \rightarrow 0^+$  if  $x \neq x_1$  and  $x \neq x_2$ , and that, for  $t > 0$ ,

$$\int_{-\infty}^{\infty} u(x, t) dx = P_1 + P_2.$$

- (c) Suppose that  $x_1, x_2, \dots, x_n$  are different points on the  $x$ -axis. Use the results from part (b) to write down the solution of the diffusion equation which has a point source at each  $x_i$  with respective strength  $P_i$ .
- (d) Explain how the solution (4.47) can be interpreted in terms of point source solutions.

**4.12** This problem considers the solution for different types of step functions in the initial condition.

- (a) Suppose the initial condition is

$$u(x, 0) = \begin{cases} u_L & \text{if } x < a, \\ u_R & \text{if } a < x. \end{cases}$$

Show that by making the change of variables  $z = x - a$ , that  $u$  satisfies the diffusion equation (in  $z$  and  $t$ ), and the initial condition resembles the one in (4.25). Use this to write down the solution, and show that after converting back to  $x$  and  $t$ ,

$$u(x, t) = u_R + \frac{1}{2}(u_L - u_R)\operatorname{erfc}\left(\frac{x - a}{2\sqrt{Dt}}\right).$$

- (b) Suppose the initial condition is

$$u(x, 0) = \begin{cases} u_L & \text{if } x < a, \\ u_M & \text{if } a < x < b, \\ u_R & \text{if } b < x. \end{cases}$$

The solution can be written as  $u = u_1 + u_2$ , where  $u_1$  and  $u_2$  are solutions of the diffusion equation, where  $u_1(x, 0)$  has a jump only at  $x = a$ , and  $u_2(x, 0)$  has a jump only at  $x = b$ . Find  $u_1(x, 0)$  and  $u_2(x, 0)$ , and then use the result in part (a) to write down the solution  $u(x, t)$ .

## Section 4.5

**4.13** Find the differential equation satisfied by the Fourier transform  $U(k, t)$ . Assume that  $-\infty < x < \infty$ , and that the solution and its derivatives go to zero as  $x \rightarrow \pm\infty$ .

- (a)  $u_t + u_x = u_{xxx}$ .
- (b)  $u_t + u_{xxx} = e^{-2x}I_{(0,\infty)}(x)$ .
- (c)  $u_t + u_x = e^{-|x|}$ .
- (d)  $u_t = u_x + 2xI_{(-1,3)}(x)$ .
- (e)  $u_t + xu_x = 0$ .

**4.14** This problem concerns calculating the Fourier transform or its inverse.

(a) Find  $f(x)$  if

$$F(k) = \frac{1}{(1 + ik)(2 + ik)}.$$

(b) Find  $f(x)$  if

$$F(k) = \frac{1}{(2 + ik)} e^{-ik}.$$

(c) Find  $F(k)$  if

$$f(x) = \begin{cases} \cosh(x) & \text{if } |x| \leq \alpha, \\ 0 & \text{otherwise.} \end{cases}$$

**4.15** This problem develops some of the basic properties of the Fourier transform. Assuming  $F(k)$  is the Fourier transform of  $f(x)$  show the following.

- (a) The Fourier transform of  $f(ax)$ , for  $a \neq 0$ , is  $F(k/a)/|a|$ .
- (b) The Fourier transform of  $f(x - a)$  is  $e^{-iak} F(k)$ .
- (c) The Fourier transform of  $f(x) \cos(ax)$  is  $\frac{1}{2} (F(k + a) + F(k - a))$ .

**4.16** Suppose the initial condition for the diffusion problem is

$$u(x, 0) = \begin{cases} u_0 & \text{if } |x| \leq h, \\ 0 & \text{otherwise.} \end{cases}$$

Show that the solution is

$$u(x, t) = \frac{1}{2} u_0 \left[ \operatorname{erf} \left( \frac{x + h}{2\sqrt{Dt}} \right) + \operatorname{erf} \left( \frac{h - x}{2\sqrt{Dt}} \right) \right],$$

where  $\operatorname{erf}()$  is the error function.

**4.17** This problem concerns the convection-diffusion equation

$$u_t = Du_{xx} - cu_x, \quad \text{for } \begin{cases} -\infty < x < \infty, \\ 0 < t, \end{cases}$$

with the initial condition  $u(x, 0) = f(x)$ . Assume  $c$  is a constant.

- (a) Using the Fourier transform, find the solution of the above problem.
- (b) Make the change of variables  $\xi = x - ct$ ,  $\tau = t$ . Letting  $v(\xi, \tau) = u(x, t)$  show that  $v$  satisfies a problem very similar to the one solved in Sect. 4.5. Use this observation, and (4.47), to write down the solution of the convection-diffusion problem.

**4.18** This problem concerns the reaction-diffusion equation

$$u_t = Du_{xx} - cu, \quad \text{for } \begin{cases} -\infty < x < \infty, \\ 0 < t, \end{cases}$$

with the initial condition  $u(x, 0) = f(x)$ . Assume  $c$  is a positive constant.

- What reaction(s) give rise to this reaction-diffusion equation?
- Using the Fourier transform, find the solution of the above problem.
- Show that the problem can also be solved by first letting  $u = ve^{at}$ , where  $a$  is a constant of your choosing, and then using (4.47).
- Suppose the  $-cu$  term in the differential equation is replaced with  $cu$ . What reaction(s) give rise to the resulting reaction-diffusion equation?

**4.19** This problem concerns solving the wave equation

$$u_{tt} = c^2 u_{xx}, \quad \text{for } \begin{cases} -\infty < x < \infty, \\ 0 < t, \end{cases}$$

with the initial conditions  $u(x, 0) = f(x)$  and  $u_t(x, 0) = 0$ . Assume  $c$  is a positive constant, and  $f(x)$  and its derivatives go to zero as  $x \rightarrow \pm\infty$ . Using the Fourier transform, find the solution of this problem.

**4.20** In this problem the inverse Fourier transform for the diffusion equation is derived from scratch.

- Show that

$$u(x, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(k) e^{ikx - Dk^2 t} dk.$$

- Setting  $H(k) = \mathcal{F}(e^{-x^2})$ , show that  $H' = -\frac{k}{2}H$ . Using the fact that  $H(0) = \sqrt{\pi}$  show that  $\mathcal{F}(e^{-x^2}) = \sqrt{\pi} e^{-k^2/4}$ . From this show that

$$\int_{-\infty}^{\infty} e^{ikq - Dk^2 t} dk = \sqrt{\frac{\pi}{Dt}} e^{-q^2/(4Dt)}.$$

- Use parts (a) and (b) to derive (4.47).

## Section 4.6

**4.21** Assume that  $c$  satisfies (4.55) for  $0 < x < L$ , and the boundary conditions are  $J = 0$  at  $x = 0$  and at  $x = L$ .

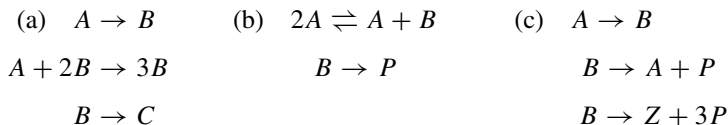
- Explain what the boundary conditions mean physically.
- Assume the initial condition is  $c(x, 0) = g(x)$ . Find the steady-state solution, and explain how it connects with your answer in part (a).

**4.22** Using the steady-state solution (4.60) to find the diffusion coefficient requires the experiment to run for almost three hours. Explain how it is possible to find  $D$  within 60 s of the start of the experiment.

**4.23** In this problem you are to write down the mathematical problem coming from the stated model. Assume the spatial interval is  $0 < x < \ell$  and  $Q = 0$ .

- (a) Fickian diffusion, with a concentration of zero at  $x = 0$ , and a flux of zero at  $x = \ell$ . Also, the initial concentration is  $c_0 x(2\ell - x)/\ell^2$ .
- (b) Drift diffusion, with a concentration of one on the left end, and a flux of  $-3$  on the right end. Also, the concentration is zero at the beginning.
- (c) Nernst-Planck diffusion, with a concentration of one on the right end, a flux of zero on the left end, and the initial concentration is  $\sin(2\pi x/\ell)$ .
- (d) Fickian diffusion, with a no flux condition at the left end, and a concentration 5 at the right end. Also, the concentration is  $5x/\ell$  at the beginning.

**4.24** The following reactions give rise to reaction diffusion equations. You are to do the following: (1) write down the kinetic equations for the reaction, find a conservation law, and determine the steady states, (2) write down the resulting reaction diffusion equations, (3) explain why the steady state you found in part (1) is a steady-state solution of the reaction diffusion equations, and (4) explain why the conservation law you found in part (1) is not necessarily a conservation law for the reaction diffusion equations. Assume  $0 < x < 1$ , and each species satisfies a no flux condition at the two endpoints.



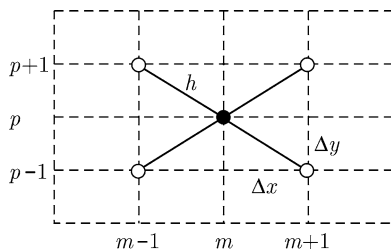
## Section 4.7

**4.25** In two dimensions the lattice need not be square, or even rectangular. This problem examines what happens in such cases.

- (a) Suppose in the lattice shown in Fig. 4.26,  $\Delta x$  and  $\Delta y$  are not equal. Assuming  $\lambda = \Delta x/\Delta y$  is fixed what is the resulting diffusion equation?
- (b) Suppose the lattice is as shown in Fig. 4.32, where  $\Delta x = \Delta y$ . In this case the particle jumps along the diagonal, with the four possibilities having equal probability. Show that one still obtains the diffusion equation in (4.79) with  $D$  given in (4.80).

**4.26** This problem explores how to derive the diffusion equation for the general random walk in the plane, as given in (4.72) and (4.73). Let  $u(x, y, t)$  be the probability that the particle is located at the spatial location  $(x, y)$  at time  $t$ .

**Fig. 4.32** Random walk for Exercise 4.25. A molecule at the black dot will move, with equal probability, to one of the hollow dots. The step length is  $h$



- (a) Suppose that at time step  $t + \Delta t$  the particle is located at  $(x, y)$ . Explain why at time  $t$  the particle was located somewhere on the circle of radius  $h$  that is centered at  $(x, y)$ .
- (b) As an approximation to the circle in part (a), distribute  $n$  points uniformly around this circle. Specifically, take the points  $(x + h \cos(j \Delta \theta), y + h \sin(j \Delta \theta))$ , where  $\Delta \theta = 2\pi/n$  and  $j = 1, 2, \dots, n$ . Explain why the probability of the particle moving from one of these  $n$  points to  $(x, y)$  is approximately  $1/n$ . From this explain why

$$u(x, y, t + \Delta t) \approx \frac{1}{n} \sum_{j=1}^n u(x + h \cos(j \Delta \theta), y + h \sin(j \Delta \theta), t).$$

- (c) Use the result from part (b) to show that for the general random walk

$$u(x, y, t + \Delta t) = \frac{1}{2\pi} \int_0^{2\pi} u(x + h \cos \theta, y + h \sin \theta, t) d\theta.$$

- (d) Derive the diffusion equation from the result in part (c) by letting  $\Delta t$  and  $h$  approach zero.

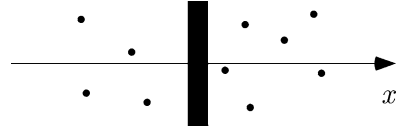
## Section 4.8

**4.27** This problem explores how to derive the Langevin equation from Newton's second law. The situation is illustrated in Fig. 4.33, which shows a rectangular object, what will be referred to as the Brownian particle, and it has mass  $M$ . There are also numerous neighboring smaller particles, each with mass  $m$ . All particles are assumed to only move along the  $x$ -axis. When one of the smaller particles, with velocity  $v$ , collides with the Brownian particle, which has velocity  $V$ , then the Brownian particle's after-collision velocity  $V'$  is

$$V' = \frac{M - m}{M + m} V + \frac{2m}{M + m} v.$$



**Fig. 4.33** Brownian particle and neighboring smaller particles used in Exercise 4.27



- (a) The total momentum and kinetic energy of the two particles are, respectively,  $mv + MV$  and  $(mv^2 + MV^2)/2$ . Assuming the values of these two quantities are conserved during the collision, derive the above formula for  $V'$ . What is the after-collision velocity  $v'$  of the smaller particle?
- (b) Suppose the Brownian particle collides with a particle with velocity  $v_1$  and then collides with a second particle with velocity  $v_2$ . If  $V_0$  and  $V_2$  are the Brownian particle's velocity before and after these two collisions, show that

$$V_2 = L^2 V_0 + \frac{2\varepsilon}{1 + \varepsilon} (Lv_1 + v_2),$$

where  $L = (1 - \varepsilon)/(1 + \varepsilon)$ , and  $\varepsilon = m/M$ .

- (c) Suppose there are successive collisions with particles with respective velocities  $v_1, v_2, \dots, v_N$ . What is the resulting formula for  $V_N$  for the Brownian particle?
- (d) Suppose that over the time interval from  $t$  to  $t + \Delta t$  that the Brownian particle collides with  $N$  smaller particles. Assuming that  $\varepsilon \ll 1$ , and using the result from part (c), show that

$$V(t + \Delta t) - V(t) \sim -2N\varepsilon V(t) + 2\varepsilon \sum_{i=1}^N v_i.$$

In deriving this, explain why it is necessary to also assume that  $N\varepsilon \ll 1$ .

- (e) Letting  $n$  denote the average number of collisions per second, then  $N \approx n\Delta t$ . Letting  $\Delta t$  approach zero, use the result from part (d) to show that

$$M \frac{dV}{dt} = -\mu V + F,$$

where  $\mu = 2mn$  and, assuming the limit exists,

$$F = \lim_{\Delta t \rightarrow 0} \frac{2m}{\Delta t} \sum_{i=1}^{n\Delta t} v_i.$$

- (f) Suppose that  $\alpha N$  particles have velocity  $v$ , and  $\beta N$  particles have velocity  $-v$ , where  $\alpha + \beta = 1$ . What is  $F$  in this case?

# Chapter 5

## Traffic Flow



### 5.1 Introduction

In this chapter we again investigate the movement of objects along a one-dimensional path, but now the motion is directed rather than random. Examples of such situations include:

- Cars moving along a highway (Fig. 5.1)
- Water flowing through a hose (Fig. 5.2)
- Astrophysical jets emitted during the formation of a star (Fig. 5.3).

Although the underlying physics of each of these is quite different, they all involve the movement of objects along what is effectively a one-dimensional pathway. We will take advantage of this when developing a mathematical model for the motion, but before doing so we must first decide on the scale we will use to characterize the motion. For example, in the last chapter we used a random walk model to study diffusion. There are particle models for fluids and traffic flow, but to start we will use a continuum description. Later in the chapter, in Sect. 5.7, a cellular automata model for traffic flow is considered and compared to the continuum version. A particle model, which involves the development and analysis of a master equation, is considered in Exercise 5.24.

### 5.2 Continuum Variables

We are assuming that the objects are numerous enough that it is not necessary to keep track of each one individually, and we can use an averaged value. In deriving the mathematical model, the objects here will be identified as cars and the path as a highway. There are a couple of reasons for using this particular example. One is that mostly everyone has experience with traffic, and is able to relate the mathematical



**Fig. 5.1** Aerial view of traffic flow (Maps 2007)



**Fig. 5.2** Flow of water molecules through, and out of, a hose

results with the real-world application. The other reason is that the theory for traffic flow is still not complete, so there are competing ideas that can be explored. However, it should be remembered that all of this material can be applied to other systems, such as the one-dimensional motion of blood cells and molecules. In fact, some of the terminology that is introduced comes from gas dynamics, because of its early use of the ideas developed here.

### 5.2.1 *Density*

The variable that will play a prominent role in our study is the traffic density  $\rho(x, t)$ . This is the number of cars per unit length, and it is instructive to consider how it might be determined experimentally. To measure  $\rho$  at  $x = x_0$ , for  $t = t_0$ , one selects



**Fig. 5.3** Artist's impression of astrophysical jets emitted during the formation of a star (ESO/L. Calçada/M. Kornmesser 2019)

a small spatial interval  $x_0 - \Delta x < x < x_0 + \Delta x$  on the highway, and then counts the number of cars within this interval (see Fig. 5.4). In this case

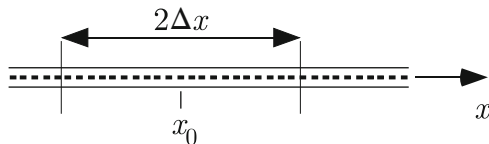
$$\rho(x_0, t_0) \approx \frac{\text{number of cars from } x_0 - \Delta x \text{ to } x_0 + \Delta x \text{ at } t = t_0}{2\Delta x}. \quad (5.1)$$

The underlying assumption here is that  $\Delta x$  is small enough that only cars in the immediate vicinity of  $x_0$  are used to determine the density at this point. At the same time,  $\Delta x$  cannot be so small that it is on the order of the length of individual cars (and the spacing between them). In the continuum viewpoint, the cars are distributed smoothly over the entire  $x$ -axis, and the value of  $\rho(x_0, t_0)$  is the limit of the right-hand side of (5.1) as  $\Delta x \rightarrow 0$ .

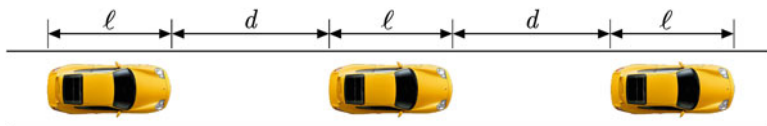
*Example (Uniform Distribution)* To illustrate how density is determined, suppose the cars all have length  $\ell$ , and they are evenly spaced a distance  $d$  apart (see Fig. 5.5). Given a sampling interval  $2\Delta x$  along the highway, then the number of cars in this interval is, approximately,  $2\Delta x/(\ell + d)$ . Inserting this into (5.1) and letting  $\Delta x \rightarrow 0$  we find that

$$\rho = \frac{1}{\ell + d}. \quad (5.2)$$

One conclusion that comes from this formula is that there is a maximum density. Because  $0 \leq d < \infty$ , then  $0 < \rho \leq \rho_M$ , where  $\rho_M = 1/\ell$ . For example, if  $\ell = 17$  ft (5.2 m) and  $d = 12$  ft (3.6 m), then, recalling  $1 \text{ mi} = 5280 \text{ ft}$ ,  $\rho = 182 \text{ cars/mi}$  (113 cars/km). With these dimensions, then the maximum density that is possible, which occurs when  $d = 0$ , is  $\rho_M = 310.6 \text{ cars/mi}$  (193 cars/km). When studying traffic flow, it is useful to know the maximum merge density  $\rho_{mg}$ , which corresponds



**Fig. 5.4** The interval along the highway used to calculate an approximate value of the density  $\rho(x_0, t_0)$ . It is also used to derive the balance law for traffic flow



**Fig. 5.5** For a uniform distribution, the cars are all the same length and are evenly spaced along the highway

to the density that occurs when the spacing is such that exactly one car fits between two cars currently on the highway. This occurs when  $d = \ell$  and for this example  $\rho_{mg} = 155.3$  cars/mi (96.5 cars/km). ■

## 5.2.2 Flux

The second variable we need is the flux  $J(x, t)$ , which has the dimensions of cars per unit time. To measure  $J$  at  $x = x_0$ , for  $t = t_0$ , one selects a small time interval  $t_0 - \Delta t < t < t_0 + \Delta t$  and counts the net number of cars that pass  $x = x_0$  during this time period. The convention is that a car moving to the right is counted as  $+1$ , while one moving to the left is counted as  $-1$ . In this case

$$J(x_0, t_0) \approx \frac{\text{net number of cars that pass } x_0 \text{ from } t = t_0 - \Delta t \text{ to } t = t_0 + \Delta t}{2\Delta t}. \quad (5.3)$$

The underlying assumption here is that  $\Delta t$  is small enough that only cars that are passing  $x_0$  at, or near,  $t = t_0$  are used to determine the flux at  $t_0$ . At the same time, from an experimental standpoint,  $\Delta t$  cannot be so small that no cars are able to pass this location during this time interval. In the continuum viewpoint we are taking the cars are distributed smoothly over the entire  $t$ -axis and the value of  $J(x_0, t_0)$  is the limit of the right-hand side of (5.3) as  $\Delta t \rightarrow 0$ .

*Example (Uniform Flow)* Returning to the previous example of uniformly distributed cars, shown in Fig. 5.5, we now add in the assumption that the cars are moving with a constant positive velocity  $v$ . In this case, the cars that start out a

distance  $2\Delta t v$  from  $x_0$  will pass  $x_0$  in the time interval from  $t_0 - \Delta t$  to  $t_0 + \Delta t$ . The corresponding number of cars is, approximately,  $2v\Delta t/(\ell + d)$ . Inserting this into (5.3), and letting  $\Delta t \rightarrow 0$ , yields

$$J = \frac{v}{\ell + d}. \quad (5.4)$$

For example, if  $\ell = 17$  ft,  $d = 51$  ft and  $v = 70$  mph, then  $J = 5435$  cars/h. Also, note that  $J = \rho v$ , which is one of the fundamental formulas in traffic flow. ■

### 5.3 Balance Law

To derive an equation for the density we will use what is known as a control volume argument. For this problem the control volume is a small region on the highway, from  $x_0 - \Delta x$  to  $x_0 + \Delta x$ . This interval is shown in Fig. 5.1. During the time period from  $t = t_0 - \Delta t$  to  $t = t_0 + \Delta t$  it is assumed that the number of cars in this interval can change only due to cars entering or leaving at the left or right ends of the interval. We are therefore assuming cars do not disappear, or pop into existence, on the highway. Actually, this could happen if we were to include an off- or onramp, but this modification will be postponed for the moment (see Exercise 5.29). As stated, our balance law for cars within the highway interval is

$$\begin{aligned} & \left\{ \text{number of cars in interval at } t = t_0 + \Delta t \right\} \\ & - \left\{ \text{number of cars in interval at } t = t_0 - \Delta t \right\} \\ & = \left\{ \text{net number of cars that cross } x_0 - \Delta x \text{ from } t_0 - \Delta t \text{ to } t_0 + \Delta t \right\} \\ & - \left\{ \text{net number of cars that cross } x_0 + \Delta x \text{ from } t_0 - \Delta t \text{ to } t_0 + \Delta t \right\}. \end{aligned}$$

Rewriting this using (5.1) and (5.3) yields

$$\begin{aligned} & 2\Delta x [\rho(x_0, t_0 + \Delta t) - \rho(x_0, t_0 - \Delta t)] \\ & = 2\Delta t [J(x_0 - \Delta x, t_0) - J(x_0 + \Delta x, t_0)]. \end{aligned}$$

Using Taylor's theorem, we have that

$$\begin{aligned}
 & 2\Delta x \left( \rho + \Delta t \rho_t + \frac{1}{2}(\Delta t)^2 \rho_{tt} + \frac{1}{6}(\Delta t)^3 \rho_{ttt} + \cdots \right. \\
 & \quad \left. - \rho - \Delta t \rho_t - \frac{1}{2}(\Delta t)^2 \rho_{tt} + \frac{1}{6}(\Delta t)^3 \rho_{ttt} + \cdots \right) \\
 & = 2\Delta t \left( J - \Delta x J_x + \frac{1}{2}(\Delta x)^2 J_{xx} - \frac{1}{6}(\Delta x)^3 J_{xxx} + \cdots \right. \\
 & \quad \left. - J - \Delta x J_x - \frac{1}{2}(\Delta x)^2 J_{xx} - \frac{1}{6}(\Delta x)^3 J_{xxx} + \cdots \right),
 \end{aligned}$$

where  $\rho$  and  $J$  are evaluated at  $(x_0, t_0)$ . Collecting the terms in the above equation,

$$\rho_t + O((\Delta t)^2) = -J_x + O((\Delta x)^2).$$

Letting  $\Delta x \rightarrow 0$  and  $\Delta t \rightarrow 0$  we conclude that

$$\frac{\partial \rho}{\partial t} = -\frac{\partial J}{\partial x}. \quad (5.5)$$

This is our balance law for motion along the  $x$ -axis. It is applicable to any continuous system in which the objects are not created or destroyed. This is why it was also obtained when deriving the model for diffusion (4.53).

### 5.3.1 Velocity Formulation

It is possible to express the balance law somewhat differently, by introducing the velocity  $v(x, t)$  of the cars on the highway. This requires care because the velocity, like the other continuum variables, is an averaged quantity. To explain how this is done, consider a small interval on the highway as shown in Fig. 5.4. One measures  $v(x_0, t_0)$  experimentally by finding the average velocity of the cars in this interval. Specifically, if there are  $n$  cars in the interval, and they have velocities  $v_1, v_2, \dots, v_n$ , then

$$v(x_0, t_0) \approx \frac{1}{n} \sum_{i=1}^n v_i.$$

In the continuum model it is assumed that the limit of this average, when letting  $\Delta x \rightarrow 0$ , exists, and its value is the velocity  $v(x_0, t_0)$ .

With the above definition, the velocity is assumed to be related to the flux through the equation

$$J = \rho v. \quad (5.6)$$

This equation was derived in the uniform distribution example discussed earlier. It is also possible to derive it for situations where the velocity is not constant (see Exercise 5.27). However, a proof for the general case is not available, and so the above formula is an assumption. Some avoid this difficulty by using (5.6) as the definition of the flux, while others use it as the definition of the velocity.

Introducing (5.6) into (5.5) gives us

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho v) = 0. \quad (5.7)$$

In solving this equation it will be assumed the initial density is known, that is,

$$\rho(x, 0) = f(x). \quad (5.8)$$

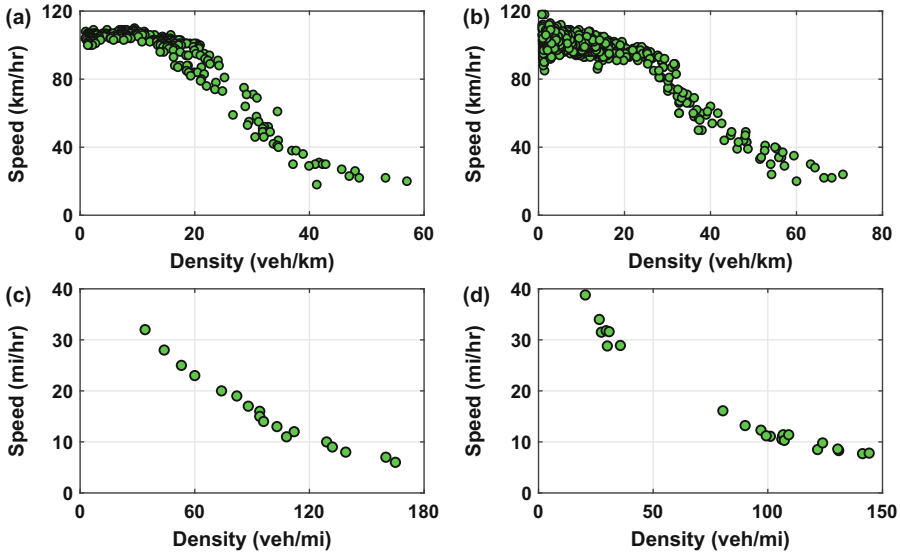
The equation in (5.7) is the mathematical model for traffic flow that we will investigate in the first part of this chapter. Those working in traffic flow refer to this as the Lighthill-Whitham-Richards (LWR) model, naming it after those who originally derived the equation (Lighthill and Whitham 1955; Richards 1956). However, the equation has wide applicability and appears under different banners. For example, in continuum mechanics it is known as the continuity equation, while in electrodynamics (5.7) is the current continuity equation, where  $\rho$  is the current density and  $J$  is the current volume. Those interested in more mathematical pursuits refer to (5.7) as a scalar conservation law.

It should be kept in mind that, as with most mathematical models, (5.7) is an approximation of the true system. Not unexpectedly, there are limitations on its applicability. As a case in point, it is questionable whether the model provides an accurate description at low densities. If the objects are few and far between, then the assumptions made in defining the density and flux are not valid. This will not stop us from using the model in such rarified regimes, but when this is done it should be understood that the continuum model provides more of a qualitative description of the motion. That said, in the regimes where it does apply, the continuum model has proven to be an exceptionally accurate, and mathematically interesting, description.

## 5.4 Constitutive Laws

Although we have derived the balance law for traffic flow, the mathematical model is incomplete. The issue is the velocity  $v$  and how it is related to the density  $\rho$ . One possibility is to investigate the physics of the problem a bit more and see if





**Fig. 5.6** The velocity as a function of the density as measured for different roadways. Shown is (a) a highway near Toronto, (b) a freeway near Amsterdam, (c) the Lincoln Tunnel, and (d) the Merritt Parkway. Data for (a) and (b) are from Rakha and Aerde (2010), and (c) and (d) are from Greenberg (1959)

there is another equation relating these variables. This is done in mechanics, and Newton's second law is used to derive a force balance equation that can be used to find the velocity. This option is not easily adaptable to the traffic flow situation so we will take a different approach and postulate how  $v$  and  $\rho$  are related based on experimental evidence. What we will be doing is specifying a constitutive law relating the velocity and density. To do this the data for several rather different roadways are shown in Fig. 5.6. The question is, what function best describes the data in this figure? The answer depends, in part, on what density and velocity intervals are of interest and what applications one has in mind. A few possible constitutive laws are discussed below.

It is worth making a couple of comments about Fig. 5.6 that are unrelated to constitutive modeling. The data in the lower two graphs were used in the original development of the continuum traffic model, while the data in the upper two graphs are typical of more modern testing. One of the striking differences between the upper and lower graphs is the amount of data shown. This is due to the development of computerized testing systems, which have been invaluable for modern scientific research. However, what is interesting is the rather tight pattern in the earlier data as compared to the scatter in the more recent results. This begs the question of whether these earlier experimentalists were more careful, or did they force the results. It makes one wonder.

### 5.4.1 Constant Velocity

The simplest assumption is that  $v$  is constant in terms of its dependence on  $\rho$ , in other words,  $v = a$ . In this case the balance law (5.7) reduces to

$$\frac{\partial \rho}{\partial t} + a \frac{\partial \rho}{\partial x} = 0. \quad (5.9)$$

This is known as the advection equation. In looking at the data in Fig. 5.6 one might conclude that assuming  $v$  is constant borders on delusional. The value of this assumption is not its realistic portrayal of traffic but, rather, what it provides in terms of insights into the type of mathematical problem that arises in traffic flow. The analysis of this problem will provide the foundation needed for solving the more difficult nonlinear problems arising from more realistic velocity functions.

### 5.4.2 Linear Velocity: Greenshields Law

The most widely used, and most well-known, constitutive laws are linear. For the traffic problem this means we assume  $v = a - b\rho$ , where  $a, b$  are constants. Those working in traffic flow refer to this as the *Greenshields model*, and the usual way this is written is

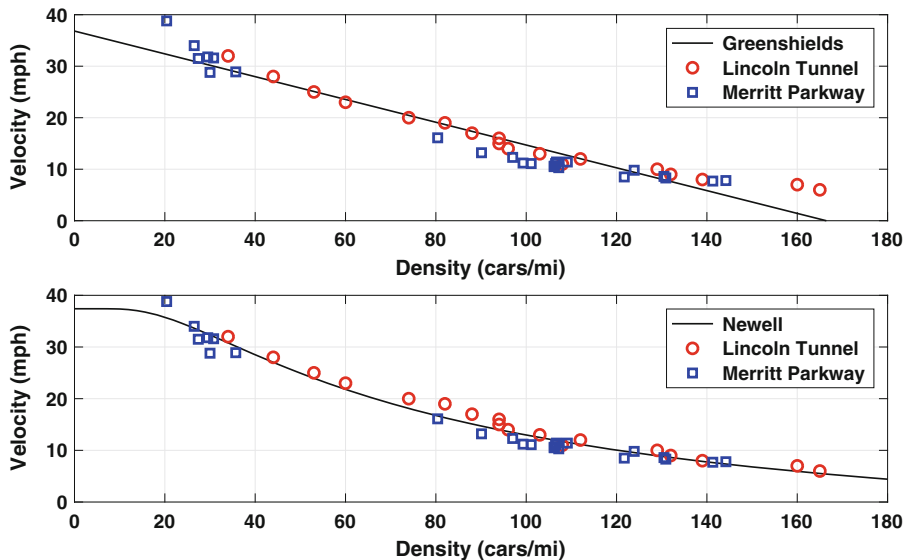
$$v = v_M \left( 1 - \frac{\rho}{\rho_M} \right), \quad (5.10)$$

where the constants  $v_M, \rho_M$  are the maximum velocity and density, respectively. The values of these constants can almost be read off the plot in Fig. 5.6. However, a more systematic way to find them is to use a least squares fit. For example, using the data for the Lincoln Tunnel and Merritt Parkway one finds that  $v_M = 36.8$  mph,  $\rho_M = 166.4$  cars/mi and the resulting function is plotted in Fig. 5.7 along with the original data. It is seen that even though this function misses the values at the extreme ends, where  $\rho = 0$  or  $\rho = 180$ , it does show the correct monotonic dependence of the velocity on density. This would seem an acceptable approximation, and the traffic flow equation (5.7) reduces to

$$\frac{\partial \rho}{\partial t} + c(\rho) \frac{\partial \rho}{\partial x} = 0, \quad (5.11)$$

where

$$c = v_M \left( 1 - \frac{2\rho}{\rho_M} \right). \quad (5.12)$$



**Fig. 5.7** Curve fit of the Greenshields law (5.10) and the Newell law (5.17) to traffic data for the Merritt Parkway and the Lincoln Tunnel

This is a nonlinear conservation equation for  $\rho$ . It can be solved analytically, but it is certainly more challenging than the linear equation in (5.9). We will return to this problem once we have worked out the constant velocity case later in this chapter.

### 5.4.3 General Velocity Formulation

It is clear from the data in Fig. 5.6 that the relationship between the velocity and density is not linear. In certain applications these differences are considered significant, and a more accurate function is needed. The general version of the constitutive law in this case has the form

$$v = F(\rho). \quad (5.13)$$

With this, the general formula for the flux is  $J = \rho F(\rho)$ . Assuming that  $F$  is a smooth function of  $\rho$ , then, using the chain rule, it follows that  $\frac{\partial}{\partial x} J = J'(\rho) \frac{\partial}{\partial x} \rho$ . The general form of the balance law (5.5) now takes the form

$$\frac{\partial \rho}{\partial t} + c(\rho) \frac{\partial \rho}{\partial x} = 0, \quad (5.14)$$

where

$$c(\rho) = J'(\rho), \quad (5.15)$$

or equivalently

$$c(\rho) = F(\rho) + \rho F'(\rho). \quad (5.16)$$

The function  $c(\rho)$  is known as the *wave velocity*, and it will play a critical role in the solution of the equation. A particular example of this function is given in (5.12), which is the wave velocity associated with the Greenshields constitutive law in (5.10).

Some thought must go into deciding what function to use for the constitutive law in (5.13). For example, there is the question of whether the resulting mathematical problem has a solution. For us, we will need that the resulting wave velocity  $c(\rho)$  is continuous and monotonic. There is also the issue of simplicity. The list of functions that might be used to describe the data in Fig. 5.6 is endless. It is for this reason that in selecting a particular function one should also consider simplicity. Given the uncertainty in the experimental data, and the approximate nature of the model, it is a waste of time to construct a function that hits every data point exactly. The problem is that the condition of simplicity, like beauty, is difficult to quantify. The linear relationship in (5.10) is an example of a simple function with two parameters. Another possibility, which is not so simple, is considered in the next example.

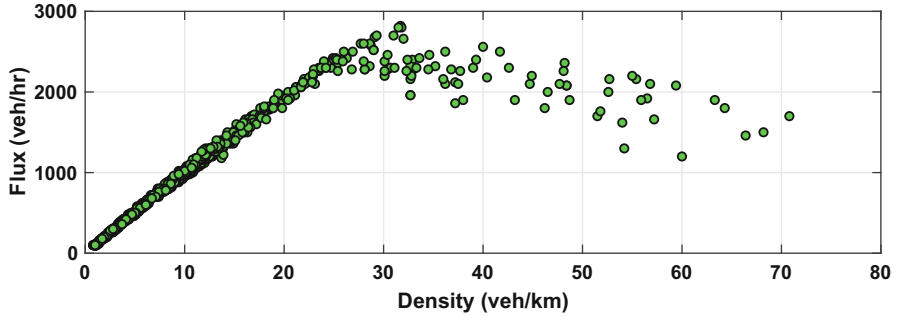
*Example (Newell Law)* A function proposed by Newell (1961) is

$$v = v_M \left( 1 - e^{-\lambda(1/\rho - 1/\rho_M)} \right). \quad (5.17)$$

Fitting this to the data for the Lincoln Tunnel and Merritt Parkway one finds that  $v_M = 37.4$  mph,  $\rho_M = 271$  cars/mi, and  $\lambda = 67.4$  mi/cars. The resulting function is plotted in Fig. 5.7 along with the original data. It is evident that it is better than Greenshields at reproducing the data and, unlike the linear law, this function contains a plateau region near  $\rho = 0$  that is seen in the Toronto and Amsterdam data in Fig. 5.6. The penalty for this improvement is that the wave velocity, given in (5.16), is

$$c = v_M \left[ 1 - \left( 1 + \frac{\lambda}{\rho} \right) e^{-\lambda(1/\rho - 1/\rho_M)} \right]. \quad (5.18)$$

One therefore has to decide if the resulting complexity in the traffic flow equation (5.14) is worth the improvement in the data fit. ■



**Fig. 5.8** The flux as a function of the density measured on a freeway in Amsterdam (Rakha and Aerde 2010)

#### 5.4.4 Flux and Velocity

Our model has three dependent variables, flux, density, and velocity. Given that the equation of motion is written in terms of density and velocity the conventional approach is to propose a constitutive law that relates these two functions. However, it is worthwhile to consider other possibilities. One alternative is to relate the flux with the density using a constitutive law, and then use the equation  $J = \rho v$  to determine the velocity. With this in mind the data in Fig. 5.6 for the freeway in Amsterdam is plotted in Fig. 5.8, giving the flux as a function of density. This is known as a fundamental diagram, and it is used extensively in developing traffic models. What is striking about this graph is that  $J$  has a well-defined dependence on  $\rho$  up to about  $\rho = 25$  after which there is considerable scatter in the data. This spread is very typical of traffic flow, and it makes formulating a constitutive law for the flux problematic. In contrast, the  $v, \rho$  plots in Fig. 5.6 show a more well-defined relationship over the entire density range, and for this reason it is more amenable to constitutive modeling.

One worthwhile conclusion that can be made from Fig. 5.8 is that the flux is concave down. In calculus it is shown that a function  $J(\rho)$  is concave down if  $J'(\rho)$  is monotone decreasing. Since  $c(\rho) = J'(\rho)$ , then we see that the wave velocity should be a monotone decreasing function of the density. It is not hard to show that both the Greenshields (5.10) and Newell (5.17) laws satisfy this requirement. In other applications where (5.14) arises, such as gas dynamics, the requirement is that the flux is concave up, with the consequence that  $c(\rho)$  needs to be monotone increasing. As we will see later, the resulting monotonicity of  $c(\rho)$  is used to guarantee that there is a solution to the problem.

#### 5.4.5 Reality Check

It is important to understand that even the most complex nonlinear expression relating the velocity and density is still, in the end, an approximation. Inevitably

certain aspects of the problem are not accounted for, and many times this is intentional because the goal of the model is to capture the essential mechanisms responsible for the phenomena being studied. This has certainly been the case with the traffic flow problem. We have not included effects of intersections, inclement weather, adverse road conditions, or myriad other things that can influence traffic flow. There is also the problem that the cars are driven by people, who make individual decisions that can have dramatic effects on the traffic pattern. As a simple example, some drivers will speed up if there is lighter traffic ahead. This implies that the velocity depends on the density gradient. This is not accounted for in our model because we are assuming that the law has the form  $v = F(\rho)$  and not  $v = F(\rho, \rho_x)$ . Some of the consequences of this extension are explored in Exercise 5.21. Generally, this sort of application is outside the scope of this textbook. However, a very humorous account of the role of human behavior, and how it affects traffic flow, can be found in Vanderbilt (2008).

A second comment that needs to be made is that the equation of motion (5.7) is general, and in terms of traffic flow can be applied to a multilane freeway or a small farm road. However, once a specific constitutive law for the velocity is introduced, then the model becomes more limited in its applicability. For example, the traffic data given in Fig. 5.6 measures the velocity on one side of the roadway (e.g., the velocity of the vehicles going east to west). This is reasonable because if both sides are counted, so the measured velocities can be either positive or negative, one could end up concluding that on average the velocity is zero at all density levels. In fact, it is not uncommon in traffic applications to have the constitutive law limited to a particular lane of traffic. For example, some roadways limit trucks to certain lanes of the roadway and this has a significant consequence for the velocity function. The point here is that the equation of motion is general but in applying it to particular problems, which requires the specification of a constitutive law, the model becomes more limited.

All of the above comments are evidence that we are studying a rich problem that has multiple research directions, and our model addresses one of them. Our objective is to understand how traffic flow behaves under the assumed conditions, and our next step is to figure out how to solve the mathematical problem we have produced.

## 5.5 Constant Velocity

To investigate the properties of the traffic flow problem we begin with the assumption that the velocity is constant. The problem takes the form

$$\frac{\partial \rho}{\partial t} + a \frac{\partial \rho}{\partial x} = 0, \quad \text{for} \quad \begin{cases} -\infty < x < \infty \\ 0 < t, \end{cases} \quad (5.19)$$

where

$$\rho(x, 0) = f(x). \quad (5.20)$$

The partial differential equation (5.19) is known as the advection equation. The solution can be found if one notes that the equation can be written as

$$\left( \frac{\partial}{\partial t} + a \frac{\partial}{\partial x} \right) \rho = 0. \quad (5.21)$$

The idea is to transform  $x, t$  to new variables  $r, s$  in such a way that the derivatives transform as

$$\frac{\partial}{\partial r} = \frac{\partial}{\partial t} + a \frac{\partial}{\partial x}. \quad (5.22)$$

If this is possible, then (5.21) becomes  $\frac{\partial \rho}{\partial r} = 0$  and this equation is very easy to solve. With this goal in mind let  $x = x(r, s), t = t(r, s)$ , in which case using the chain rule the  $r$ -derivative transforms as

$$\frac{\partial}{\partial r} = \frac{\partial x}{\partial r} \frac{\partial}{\partial x} + \frac{\partial t}{\partial r} \frac{\partial}{\partial t}. \quad (5.23)$$

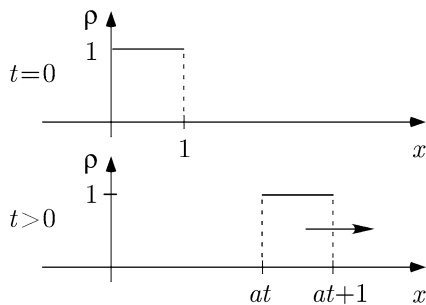
Comparing this with (5.21), we require  $\frac{\partial x}{\partial r} = a$  and  $\frac{\partial t}{\partial r} = 1$ . Integrating these equations yields  $x = ar + q(s)$  and  $t = r + p(s)$ . To determine the  $s$  dependence recall that the initial condition specifies the solution along the  $x$ -axis. To make it easy to apply the initial condition we will ask that the  $x$ -axis ( $t = 0$ ) maps onto the  $s$ -axis ( $r = 0$ ). In other words,  $r = 0$  implies that  $t = 0$  and  $x = s$ . Setting  $r = 0$  and  $t = 0$  we conclude  $q(s) = s$  and  $p(s) = 0$ , and so, the change of variable we are looking for is

$$x = ar + s, \quad t = r. \quad (5.24)$$

Inverting this transformation one finds that  $r = t$  and  $s = x - at$ . We are now able to write (5.19) as  $\frac{\partial \rho}{\partial r} = 0$ , which means  $\rho = \rho(s) = \rho(x - at)$ . With the initial condition we therefore conclude that the solution of the problem is

$$\rho(x, t) = f(x - at). \quad (5.25)$$

Before making general conclusions about this solution we consider an example. This is worked out twice, first as a mathematical problem, and then as a problem in traffic flow.



**Fig. 5.9** Solution of the advection equation (5.19). The top figure is the initial condition, as given in (5.29). The bottom figure is the solution at a later time, as given in (5.25)

*Example (Mathematical Version)* Suppose the initial condition is the square bump shown in Fig. 5.9. In mathematical terms,

$$f(x) = \begin{cases} 1 & \text{if } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases} \quad (5.26)$$

From (5.25) the solution is

$$\rho(x, t) = \begin{cases} 1 & \text{if } 0 < x - at < 1, \\ 0 & \text{otherwise,} \end{cases}$$

or equivalently,

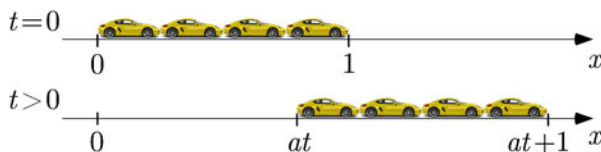
$$\rho(x, t) = \begin{cases} 1 & \text{if } at < x < 1 + at, \\ 0 & \text{otherwise.} \end{cases} \quad (5.27)$$

A typical solution profile is also shown in Fig. 5.9 in the case of when  $a > 0$ . It is apparent that at any given time  $t$ , the solution is simply the original square bump that has moved over to occupy the interval  $at \leq x \leq 1 + at$ . ■

*Example (Traffic Version)* The previous example can be restated in physical terms. The initial condition (5.29) can be interpreted as having a group of cars that are uniformly spaced as shown in Fig. 5.10. Also, because the cars all travel with a constant velocity  $a$ , they will move as a unit. So, at any given time  $t$ , the group of cars will occupy the interval  $at < x < at + 1$ . Because they are traveling at the same velocity, the spacing of the cars has not changed, and therefore the density in this interval is the same as it was at  $t = 0$ . This is the same result as obtained in the solution (5.27). ■

The attractive aspect of the traffic version is that the solution is easy to understand, and it is obtained without having to solve a partial differential equation. Unfortunately, for more realistic problems, where the velocity depends on the





**Fig. 5.10** A uniformly spaced group of cars moves with constant velocity  $a$  along the  $x$ -axis

density, the traffic version loses this advantage and the continuum problem becomes the easier one to solve.

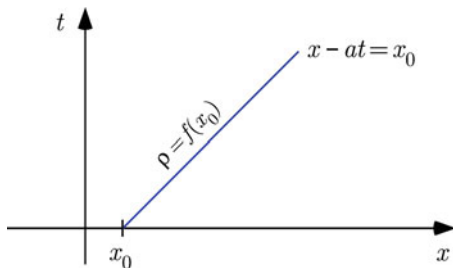
From the above examples, and from the general formula in (5.25), we conclude that the solution is a traveling wave. The wave travels in only one direction, and for this reason (5.19) is sometimes called a one-way wave equation. In the case of when  $a > 0$  the wave moves to the right with speed  $a$ . What is significant is that it moves at the same velocity as the vehicles, which, if you recall, is  $v = a$ . It might seem obvious that the wave moves with the vehicle velocity because, after all, the vehicles are responsible for the wave in the first place. However, the answer is not so simple. For example, the waves generated at sporting events by the fans in the audience are obtained not by the fans running around the stadium but, rather, by them periodically standing up and sitting down. Similarly, in heavy traffic if a car's taillights come on you will likely see a wave of taillights come on in the cars that follow. Not only is the wave of taillights not moving with the car's velocity, it is actually moving in the opposite direction. So, the connection between the motion of the constituents and the velocity of the wave requires some consideration. We will return to this point later when solving the problem of nonconstant velocity.

Another observation coming from the above example is that the shape and amplitude do not change as the wave travels along the  $x$ -axis. This is in marked contrast to the diffusion equation, where the corners or jumps in the initial condition are immediately smoothed out (see Fig. 4.17). Because of this, one might question whether (5.27) is actually a solution since  $\rho_x$  is not defined at the jumps located at  $x = at, 1 + at$ . The short answer is that because there are only a finite number of jumps, everything is fine. What is necessary is to introduce the concept of a weak solution, and the interested reader is referred to Evans (2010) for an extended discussion of this subject. A slightly different approach to justifying the jumps, and understanding some of the difficulties of defining a continuum variable at a jump, are explored in Exercise 5.28.

### 5.5.1 Characteristics

There is another way to look at this solution that will prove to be particularly worthwhile. It is based on the observation that, from the formula  $\rho(x, t) = f(x - at)$ , if we hold  $x - at$  fixed, then the solution is constant. In other words,

**Fig. 5.11** The characteristics for (5.19) are the straight lines  $x - at = x_0$ . Along each line the solution is constant



if  $x - at = x_0$ , then  $\rho = f(x_0)$  along this line (see Fig. 5.11). These lines are called *characteristics* for the equation, and the method we used to find the solution is called the method of characteristics. This result is important enough that it should be restated in a more pronounced format.

**Fundamental Property of Characteristics (linear case).** *The solution of  $\partial_t \rho + a \partial_x \rho = 0$  is constant along any line of the form  $x - at = \text{constant}$ .*

The method of characteristics uses this statement to solve the problem. You might wonder why this is being discussed since, after-all, we have already found the solution. The reason is that the nonlinear problem can also be solved using characteristics, even though a simple solution as in (5.25) is not available. The linear problem provides an easy introduction to using the method of characteristics, and it does not involve some of the complications that arise with the nonlinear problem.

*Example (Red Light–Green Light)* The objective is to use the characteristics to solve

$$\frac{\partial \rho}{\partial t} + 3 \frac{\partial \rho}{\partial x} = 0, \quad (5.28)$$

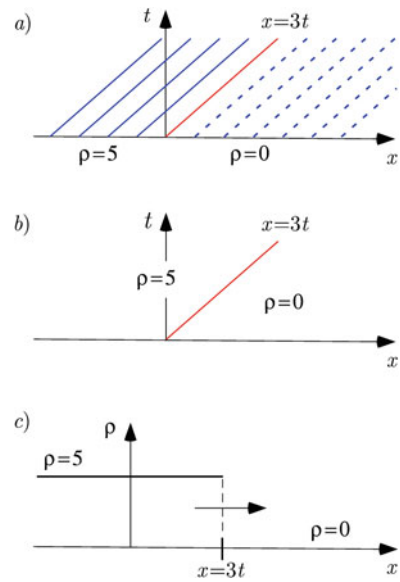
with the initial condition

$$\rho(x, 0) = \begin{cases} 5 & \text{if } x \leq 0 \\ 0 & \text{if } x > 0. \end{cases} \quad (5.29)$$

This function can be thought of as arising when cars are waiting at a stop light located at  $x = 0$ . At  $t = 0$  the light turns green, and the movement of the cars is then determined from the solution of (5.28). For this reason it is often referred to as the red light–green light problem.

The characteristics for this problem are the lines  $x - 3t = x_0$  and they are illustrated in Fig. 5.12a. Because of where the characteristics intersect the  $x$ -axis, the solution in the region covered by the solid lines is  $\rho = 5$ , while along the dashed lines the solution is  $\rho = 0$ . The characteristic that separates these two regions is the one that starts at the jump in the initial condition (5.29). Namely, it is the line  $x = 3t$ , which is shown in red in Fig. 5.12. The resulting solution is shown in Fig. 5.12b, and

**Fig. 5.12** (a) The characteristics of the red light–green light example; (b) the resulting solution in the  $x, t$ -plane; and (c) the solution in the  $x, \rho$ -plane



the corresponding formula is

$$\rho(x, t) = \begin{cases} 5 & \text{if } x \leq 3t \\ 0 & \text{if } x > 3t. \end{cases} \quad (5.30)$$

A somewhat more traditional view of the solution is given in Fig. 5.12c, where it is apparent that the solution consists of a wave that moves with speed 3. ■

*Example (Finite Length Highways)* Up to this point our highways have been infinitely long. In the real world this is rather rare, and so in this example we consider how to use the method of characteristics when the road has finite length. Specifically, we consider solving the equation

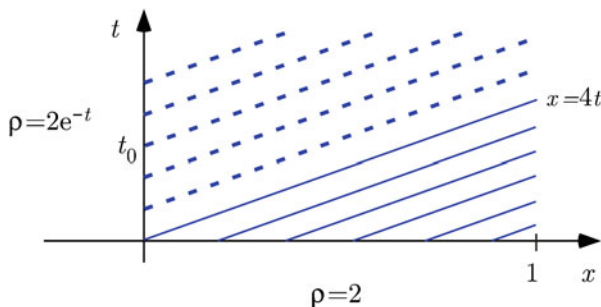
$$\frac{\partial \rho}{\partial t} + 4 \frac{\partial \rho}{\partial x} = 0, \quad \text{for } \begin{cases} 0 < x < 1 \\ 0 < t, \end{cases} \quad (5.31)$$

along with the initial condition

$$\rho(x, 0) = 2, \quad (5.32)$$

and the boundary condition

$$\rho(0, t) = 2e^{-t}. \quad (5.33)$$



**Fig. 5.13** Characteristics used in solving the traffic flow problem over a finite interval

The characteristics for this problem are the lines  $x - 4t = x_0$ , and they are illustrated in Fig. 5.13. The analysis naturally separates into two components.

**Solid Lines:** In the region containing the characteristics that are solid lines, the solution is determined by the initial condition. Because  $\rho(x, 0) = 2$ , it follows that in this region the solution is  $\rho(x, t) = 2$ .

**Dashed Lines:** To find the solution in the region where the characteristics are dashed lines, consider the characteristic shown in Fig. 5.13 that has  $t$ -intercept  $t_0$ . The equation of this characteristic is  $x - 4t = -4t_0$ . Because the solution is constant along this line, and we are told that  $\rho(0, t_0) = 2e^{-t_0}$ , then it follows that along this characteristic  $\rho(x, t) = 2e^{-t_0}$ . Since  $t_0 = t - x/4$ , then the solution can be written as  $\rho(x, t) = 2e^{-t+x/4}$ .

Putting this information together, the solution is

$$\rho(x, t) = \begin{cases} 2 & \text{if } 0 \leq t \leq x/4, \\ 2e^{-t+\frac{1}{4}x} & \text{if } x/4 < t. \end{cases}$$

■

It is worth noting that the boundary condition in the last example was given at  $x = 0$  and not at  $x = 1$ . The reason has to do with the characteristics. In Fig. 5.13, to determine the solution, we need the initial and boundary conditions to provide us with enough information to be able to determine the value of  $\rho$  on each characteristic. With the initial condition, we can determine  $\rho$  along the solid lines. Attempting to also prescribe the solution at  $x = 1$  would almost certainly lead to a conflicting requirement, with the result that there is no solution to the problem. For the same reason, if a boundary condition is prescribed at  $x = 1$ , then one would not also prescribe an initial condition or a boundary condition at  $x = 0$ . This idea is explored further in Exercise 5.8.

## 5.6 Density Dependent Velocity

The linear wave equation studied in the previous section is a valuable source of information about some of the more basic properties of the solution. The fact is, however, the assumption that the velocity is independent of the density is not correct for traffic flow. This is evident in the data given in Fig. 5.6. Precisely what constitutive law is used will be left unspecified for the moment other than to assume  $v = F(\rho)$ , where  $F(\rho)$  is smooth. As shown in Sect. 5.4.3, the traffic flow equation takes the form

$$\frac{\partial \rho}{\partial t} + c(\rho) \frac{\partial \rho}{\partial x} = 0, \quad (5.34)$$

where the *wave velocity* is

$$c(\rho) = F(\rho) + \rho F'(\rho). \quad (5.35)$$

Written this way the equation resembles the constant velocity version in (5.19) we studied earlier. One significant difference is that the wave velocity can depend on the unknown  $\rho$ , and if this happens, then (5.34) is nonlinear.

*Example (Greenshields Law)* According to the Greenshields constitutive law,

$$v = v_M \left( 1 - \frac{\rho}{\rho_M} \right). \quad (5.36)$$

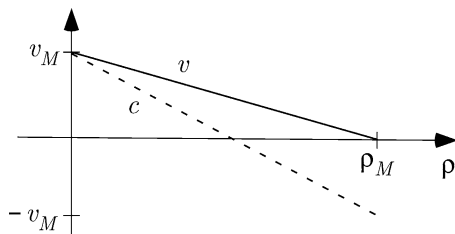
In this case, from (5.35), the wave velocity is

$$c = v_M \left( 1 - \frac{2\rho}{\rho_M} \right). \quad (5.37)$$

These functions are sketched in Fig. 5.14. It is seen that  $c \leq v$ , and, interestingly,  $c$  can be either positive or negative (even though  $v$  is nonnegative). Specifically, the wave velocity is positive for lighter traffic ( $0 \leq \rho < \rho_M/2$ ), and it is negative in heavier traffic ( $\rho_M/2 < \rho \leq \rho_M$ ). ■

The nonlinear traffic flow equation (5.34) is very general, and a couple of restrictions are needed to help guarantee that there is a solution. One is that whatever

**Fig. 5.14** The velocity  $v$ , and wave velocity  $c$ , when using the Greenshields law



function  $c(\rho)$  is used in this equation, it is a smooth function of  $\rho$ . The second condition is that  $c'(\rho) \neq 0$ , which was first mentioned in Sect. 5.4.4. Even with these assumptions, as will be seen in Sect. 5.6.5, there is a significant issue with obtaining a unique solution. Determining how to resolve this issue will give rise to some interesting questions connecting the mathematical and physical problem we are studying.

### 5.6.1 Small Disturbance Approximation

One method for studying nonlinear wave problems is based on a small disturbance approximation. The basic idea is that a particular solution has been determined. This is usually a steady-state solution, and it is very common that it is a constant. What is investigated is how small perturbations of this particular solution behave. To explain what this entails note that a constant function  $\rho = \rho_0$  is a solution of the traffic flow equation (5.34). So, suppose that the traffic is flowing along smoothly with a uniform density  $\rho = \rho_0$  and then one or more of the cars change speed slightly and cause a small perturbation in the density. For example if someone applies their brakes, then the immediate effect will be to reduce the density in front of their car and to increase the density right behind them. A function that mimics this change in the density is shown in Fig. 5.15.

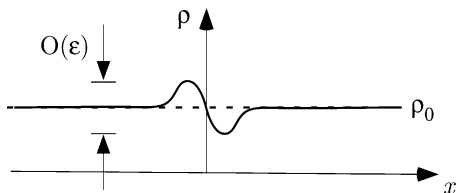
To analyze this situation we will assume the disturbance occurs at  $t = 0$ . The initial condition that corresponds to this is

$$\rho(x, 0) = \rho_0 + \varepsilon g(x). \quad (5.38)$$

The specific form of the function  $g(x)$  is not important but we will illustrate the analysis using the example in Fig. 5.15. Due to the initial condition the appropriate expansion for the solution is  $\rho \sim \rho_0 + \varepsilon \rho_1(x, t) + \dots$ . In this case, using Taylor's theorem,

$$\begin{aligned} c(\rho) &\sim c(\rho_0 + \varepsilon \rho_1 + \dots) \\ &\sim c(\rho_0) + (\varepsilon \rho_1 + \dots) c'(\rho_0) + \frac{1}{2} (\varepsilon \rho_1 + \dots)^2 c''(\rho_0) + \dots \\ &\sim c(\rho_0) + \varepsilon \rho_1 c'(\rho_0) + \dots \end{aligned}$$

**Fig. 5.15** Small disturbance imposed onto a constant density solution at  $t = 0$ . The resulting initial condition is given in (5.38)



The equation of motion (5.34) takes the form

$$\varepsilon \frac{\partial \rho_1}{\partial t} + \cdots + [c(\rho_0) + \varepsilon \rho_1 c'(\rho_0) + \cdots] \left( \varepsilon \frac{\partial \rho_1}{\partial x} + \cdots \right) = 0, \quad (5.39)$$

where, from (5.38),

$$\rho_0 + \varepsilon \rho_1(x, 0) + \cdots = \rho_0 + \varepsilon g(x). \quad (5.40)$$

Setting  $c_0 = c(\rho_0)$ , then the  $O(\varepsilon)$  problem is

$$\frac{\partial \rho_1}{\partial t} + c_0 \frac{\partial \rho_1}{\partial x} = 0, \quad (5.41)$$

where  $\rho_1(x, 0) = g(x)$ . This is known as the small disturbance equation for the problem and in this case it is a linear wave equation. Using (5.25), the solution is  $\rho_1(x, t) = g(x - c_0 t)$ . Therefore, the two term small disturbance approximation of the solution is

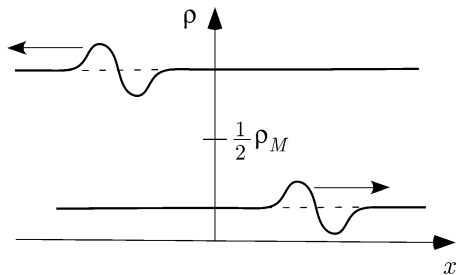
$$\rho(x, t) \sim \rho_0 + \varepsilon g(x - c_0 t). \quad (5.42)$$

It is clear from this that the initial disturbance propagates as a traveling wave, with velocity  $c_0$ . We will explore some of the consequences of this in the next example, but it is first necessary to comment on the accuracy of this approximation. If you compare (5.42) with, say, the numerical solution it is found that as time passes the approximation becomes less accurate. This is due to a slow change in the solution that is not accounted for in (5.42), and which over time starts to affect its accuracy. It is possible to use multiple scales, as described in Sect. 2.7, to improve the approximation. However, as will be shown in the next section, it is actually possible to find the exact solution using the method of characteristics.

*Example (Phantom Traffic Jams)* To investigate the properties of (5.42) we will use the Greenshields constitutive law. The wave velocity is given in (5.37) and it is shown in Fig. 5.14. Our conclusion is that in light traffic, where  $c > 0$ , the disturbance moves forward, and in heavy traffic, where  $c < 0$ , the disturbance moves backward. Given that  $c \leq v$ , the disturbance does not move faster than the flow of traffic. In other words, whoever was responsible for generating this disturbance would see it move backward relative to their position, but someone watching from an overpass would see it move forward in light traffic and move backward in heavy traffic. The one exception to this last statement is if the traffic density is  $\rho_M/2$ , in which case the disturbance would stay in the region where it was generated. Another point to make here is that, unlike the constant velocity example, the wave propagates at a velocity that is different from the velocity of the vehicles that form the system.

The solution obtained using a small disturbance approximation provides an explanation of one of the mysteries of driving called the phantom traffic jam. This

**Fig. 5.16** Disturbances move to the right if  $\rho_0 < \frac{1}{2}\rho_M$  and move toward the left if  $\frac{1}{2}\rho_M < \rho_0$ . The wave velocity in both cases is  $c_0 = c(\rho_0)$



is the situation when there is no visible reason for a traffic slowdown, as there is no accident, construction, etc. As shown in Fig. 5.16 some earlier perturbation in the traffic can result in a density wave propagating backwards along the highway. A driver who enters this region will see no apparent reason for its existence and once through the disturbance will return to the uniform flow they had earlier. One cause of such situations is weaving. In heavier traffic drivers who change lanes frequently cause the drivers behind them to slowdown or brake to leave room between them and the lane changer. This produces a small disturbance and this propagates along the highway behind the originators of this situation. ■

### 5.6.2 Method of Characteristics

As it turns out, the method of characteristics we developed to solve the constant velocity problem can be adapted so it also works on the nonlinear equation (5.34). In the constant velocity case, we found that the solution is constant along curves of the form  $x = x_0 + at$ . So, in a similar manner we will investigate if it is possible to find curves  $x = X(t)$  on which the solution of (5.34) is constant. What we are looking for are curves with the property that  $\frac{d}{dt}\rho(X(t), t) = 0$ . Expanding this using the chain rule it follows that we need to select  $X(t)$  in such a way that

$$\rho_t + X'(t)\rho_x = 0. \quad (5.43)$$

To find a function  $X(t)$  that works in this equation, recall that  $\rho$  satisfies the traffic flow equation

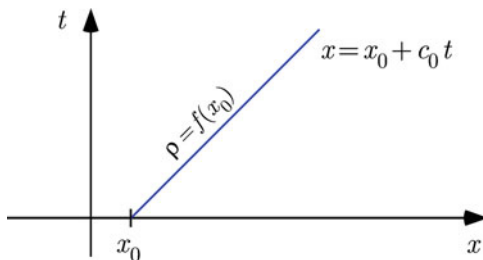
$$\rho_t + c(\rho)\rho_x = 0. \quad (5.44)$$

Comparing this with (5.43) it is evident that  $X(t)$  should be selected so that

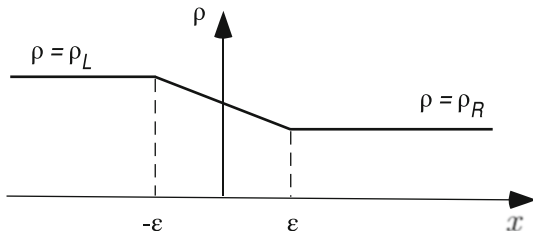
$$X'(t) = c(\rho). \quad (5.45)$$



**Fig. 5.17** The solution of (5.44) is constant along a characteristic curve  $x = x_0 + c_0 t$ , where  $c_0 = c(\rho_0)$  and  $\rho_0 = f(x_0)$



**Fig. 5.18** Initial density  $\rho(x, 0)$  for the modified red light–green light problem. Note that the slower cars are on the left



Before integrating to find the function  $X(t)$ , remember that  $\rho$  is constant along the curve. Consequently, if the curve begins at  $x = x_0$ , then at any point along the curve we have  $\rho = \rho_0$  where  $\rho_0 = f(x_0)$  (see Fig. 5.17). Introducing this into (5.45), and integrating, we obtain  $X = x_0 + c(\rho_0)t$ . This gives us the next result.

**Fundamental Property of Characteristics (general case):** *The solution of  $\partial_t \rho + c(\rho)\partial_x \rho = 0$  is constant along characteristic lines of the form  $x = x_0 + c(\rho_0)t$ , where  $\rho_0 = f(x_0)$ . In particular, along this characteristic the solution is  $\rho = \rho_0$ .*

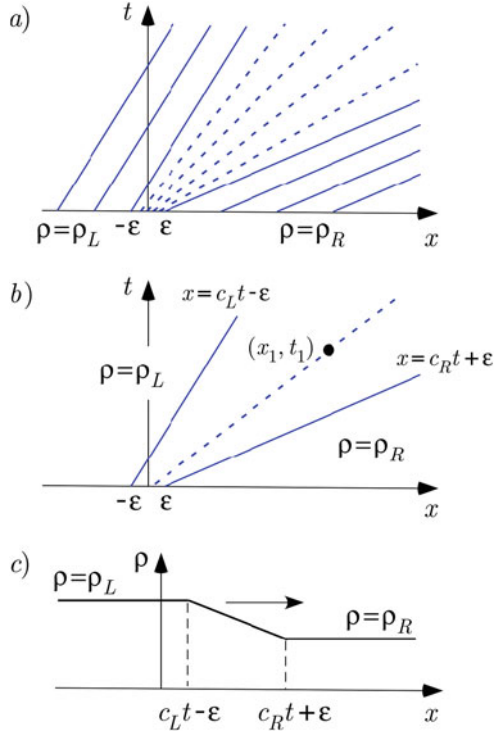
This is written for the situation shown in Fig. 5.17, and it assumes that the characteristic curves do not intersect. What happens when intersections occur is explained in Sect. 5.6.4.

It might seem odd that the characteristics for a nonlinear equation turn out to be linear. However, the nonlinearity does have an effect as it determines the slope of the characteristics and, as we will see, this has major consequences on the solution. We will postpone analyzing such difficulties until after we have more experience using characteristics when all goes according to plan.

*Example (Modified Red Light–Green Light)* To use the above solution for traffic flow we consider a modified version of the red light–green light problem. The principal modification is that there is no jump in the initial condition, and instead there is a small interval over which the change occurs. The specific initial condition to be considered is shown in Fig. 5.18, and the corresponding formula is

$$\rho(x, 0) = \begin{cases} \rho_L & \text{if } x \leq -\varepsilon \\ \rho_L + \frac{\rho_R - \rho_L}{2\varepsilon}(x + \varepsilon) & \text{if } -\varepsilon < x < \varepsilon \\ \rho_R & \text{if } \varepsilon \leq x. \end{cases} \quad (5.46)$$

**Fig. 5.19** The solution of the modified red light–green light problem as obtained using the characteristics. In this figure it is assumed that  $c_L$  and  $c_R$  are positive (see Exercise 5.8)



A significant assumption that is being made here is that the faster cars are on the right, and this means  $\rho_L > \rho_R$ . We also need to be specific about what constitutive law is being used for the velocity, and in what follows we use the Greenshields law. Consequently,  $v = v_M(1 - \rho/\rho_M)$  and the wave velocity is

$$c(\rho) = v_M \left( 1 - \frac{2\rho}{\rho_M} \right). \quad (5.47)$$

To sketch the characteristics, we consider what happens for different starting positions  $x_0$  on the  $x$ -axis.

$x_0 \leq -\epsilon$ : In this case,  $\rho(x_0, 0) = \rho_L$ . This means that the characteristics in this region all have the same slope, and this is shown in Fig. 5.19a. Given that the solution is constant along each of these lines it follows that  $\rho = \rho_L$  in the region of the  $x, t$ -plane to the left of the characteristic  $x = -\epsilon + c_L t$ , where  $c_L = c(\rho_L)$ . This is shown in Fig. 5.19b.

$x_0 \geq \epsilon$ : In this case,  $\rho(x_0, 0) = \rho_R$ . Because  $\rho_L > \rho_R$ , then  $c_L < c_R$ , where  $c_R = c(\rho_R)$ . The resulting characteristics are parallel and are shown in Fig. 5.19b. The solution is constant along each of these lines, and so it follows that  $\rho = \rho_R$  in the region of the  $x, t$ -plane to the right of the characteristic  $x = \epsilon + c_R t$ .

$-\varepsilon < x_0 < \varepsilon$ : The characteristics connected to these points are the dashed lines in Fig. 5.19a. To find the solution at a point  $(x_1, t_1)$  in this region, as illustrated in Fig. 5.19b, we need to find the characteristic that passes through this point. This requires finding  $x_0$ . Now, the general formula for a characteristic is  $x = x_0 + c_0 t$ , and so it is required that  $x_1 = x_0 + c_0 t_1$ . Given  $x_0$ , then  $c_0$  is determined using (5.47) and (5.46), which gives us

$$c_0 = v_M \left[ 1 - \frac{2}{\rho_M} \left( \rho_L + \frac{\rho_R - \rho_L}{2\varepsilon} (x_0 + \varepsilon) \right) \right].$$

Substituting this into the equation  $x_1 = x_0 + c_0 t_1$  and then solving for  $x_0$  one finds that

$$x_0 = \frac{x_1 - t_1(c_L + c_R)/2}{1 + t_1(c_R - c_L)/(2\varepsilon)}.$$

With this, and the initial condition in (5.46), the density is

$$\begin{aligned} \rho(x_1, t_1) &= \rho(x_0, 0) \\ &= \rho_L + (\rho_R - \rho_L) \frac{x_1 + \varepsilon - c_L t_1}{2\varepsilon + (c_R - c_L) t_1}. \end{aligned} \quad (5.48)$$

Based on the above discussion, the formula for the solution is

$$\rho(x, t) = \begin{cases} \rho_L & \text{if } x \leq c_L t - \varepsilon \\ \rho_L + (\rho_R - \rho_L) \frac{x + \varepsilon - c_L t}{2\varepsilon + (c_R - c_L) t} & \text{if } c_L t - \varepsilon < x < c_R t + \varepsilon \\ \rho_R & \text{if } c_R t + \varepsilon \leq x, \end{cases} \quad (5.49)$$

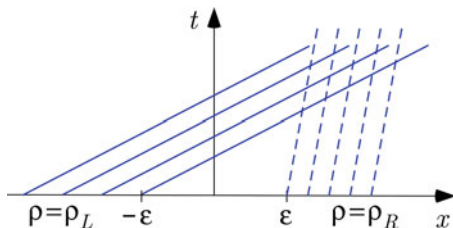
where  $c_L = c(\rho_L)$  and  $c_R = c(\rho_R)$ . According to this, between the two constant states the density varies linearly, just as it did in the initial condition.

There is nothing unusual in the solution (5.49) as it shows the expected result that the faster cars in the front gradually separate from the slower cars in the back. This is illustrated in Fig. 5.19c. What is not so obvious is where the transition points are located. As shown in the derivation, these are determined by the wave velocity  $c$ , and not by the velocity  $v$  of the cars. ■

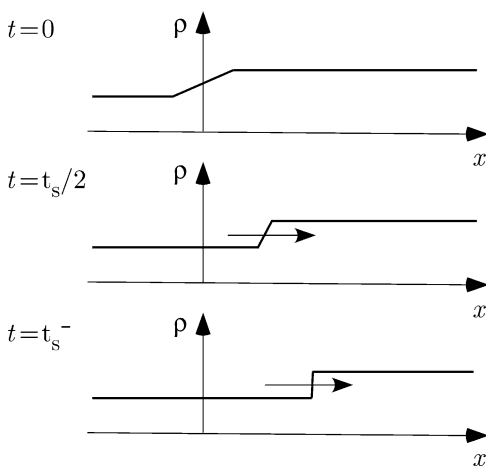
A natural, and somewhat mischievous, question to ask is, what happens if the faster cars are put in the back and the slower ones in the front? Much of what was done in the previous example can be used to answer this, and this brings us to the next example.

*Example* Suppose the piecewise linear initial condition (5.46) is used, but  $\rho_L < \rho_R$ . This means that the faster cars are in the back. Clearly, some complication is going to arise when the faster cars overtake the slower ones. To investigate what this might be, the characteristics associated with the  $x_0 \leq -\varepsilon$ , and the  $x_0 \geq \varepsilon$ , intervals are

**Fig. 5.20** The characteristics when the faster cars are placed in the back

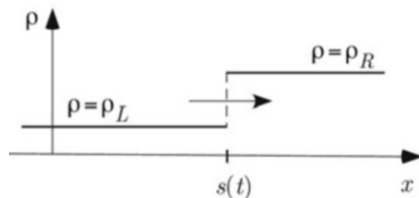


**Fig. 5.21** The solution when the faster cars are in the back. In this case, the wave steepens as it moves rightward and eventually contains a jump between the two groups of cars



shown in Fig. 5.20. Much of what was stated earlier still holds. Namely, in the region with the solid blue characteristics the solution is  $\rho = \rho_L$ , and in the region with the dashed blue characteristics  $\rho = \rho_R$ . The exception to this statement is where the two sets of characteristics overlap. In this region, the information coming from the characteristics conflicts. It is possible to get an idea of what is happening in the overlap region by looking at the solution right before the characteristics first intersect. So, let  $t = t_s$  be the time when the characteristic that starts at  $x_0 = -\varepsilon$ , and the one that starts at  $x_0 = \varepsilon$ , intersect. The solution at the start, at an intermediate time ( $t = t_s/2$ ), and right before the faster cars reach the slower group in the front ( $t = t_s^-$ ) are shown in Fig. 5.21. In drawing this figure it is assumed that  $c_L$  and  $c_R$  are positive. What is seen is that the transition region between  $\rho = \rho_L$  and  $\rho = \rho_R$  is shrinking, causing the wave to steepen as it moves rightward. It appears that when the transition region vanishes at  $t = t_s$ , the solution contains a jump that presumably continues to move rightward. This is, indeed, what happens, and the jump is known as a shock wave. To put this on a more firm mathematical footing, we need to investigate the requirements imposed on a solution with a jump discontinuity, and this is considered next. After that we will return to this problem and determine the solution. ■

**Fig. 5.22** A jump discontinuity in the solution, located at  $x = s(t)$



### 5.6.3 Rankine-Hugoniot Condition

As is evident in Fig. 5.21, the nonlinear traffic flow equation has a propensity to evolve into a function with one or more jump discontinuities that move along the  $x$ -axis. We studied such a solution with the red light-green light problem for the linear equation, and the result is shown in Fig. 5.12. The nonlinear equation is a different animal, and we are going to have to be a bit more careful any time a jump is present. To investigate what happens, suppose we have a situation as shown in Fig. 5.22, which consists of a jump that is located at  $x = s(t)$ . Given that  $x$ -derivatives are not defined at such points we will reformulate the problem by integrating over a small spatial interval,  $s - \varepsilon \leq x \leq s + \varepsilon$ , around the jump. So, integrating  $\rho_t + J_x = 0$  and remembering that the density is constant on either side of the jump we obtain

$$\int_{s-\varepsilon}^{s+\varepsilon} \rho_t dx + J(\rho_R) - J(\rho_L) = 0. \quad (5.50)$$

From the Fundamental Theorem of Calculus recall that

$$\frac{d}{dt} \int_{s-\varepsilon}^{s+\varepsilon} \rho dx = \int_{s-\varepsilon}^{s+\varepsilon} \rho_t dx + s'(t)\rho|_{x=s+\varepsilon} - s'(t)\rho|_{x=s-\varepsilon}.$$

From this, and (5.50), it follows that

$$\frac{d}{dt} \int_{s-\varepsilon}^{s+\varepsilon} \rho dx - \rho_R s'(t) + \rho_L s'(t) + J(\rho_R) - J(\rho_L) = 0. \quad (5.51)$$

Now, using the piecewise constant nature of the density

$$\begin{aligned} \int_{s-\varepsilon}^{s+\varepsilon} \rho dx &= \int_{s-\varepsilon}^s \rho dx + \int_s^{s+\varepsilon} \rho dx \\ &= \varepsilon(\rho_L + \rho_R), \end{aligned}$$

and so

$$\frac{d}{dt} \int_{s-\varepsilon}^{s+\varepsilon} \rho dx = 0.$$

It follows from (5.51) that

$$s'(t) = \frac{J(\rho_R) - J(\rho_L)}{\rho_R - \rho_L}. \quad (5.52)$$

This equation is known as the *Rankine-Hugoniot condition* (the flux version) and it determines the velocity of a jump discontinuity in the solution.

It is useful to express (5.52) in terms of the wave velocity function  $c$ . Recalling that  $c = J'(\rho)$ , and  $J(0) = 0$ , then

$$J(\rho) = \int_0^\rho c(r)dr. \quad (5.53)$$

With this, the *Rankine-Hugoniot condition* (the wave velocity version) takes the form

$$s'(t) = \frac{1}{\rho_R - \rho_L} \int_{\rho_L}^{\rho_R} c(\rho)d\rho. \quad (5.54)$$

This is an interesting result as it shows that any jump in the solution travels at the wave velocity averaged over the given density interval.

There are two types of jumps, and they are determined by what happens to the velocity  $v$  at the jump. If  $\rho$  has a jump discontinuity at  $x = s(t)$ , but  $v$  is continuous at  $x = s(t)$ , then the jump is called a *contact discontinuity*. An example is the red light-green light solution shown in Fig. 5.12. The velocity is constant, hence it is continuous no matter where the jumps occur. Note that because  $v = a$  and  $J = \rho v$ , then the Rankine-Hugoniot condition (5.52) reduces to  $s' = a$ . In other words, the jumps move with the given constant velocity, and this is what was determined in Fig. 5.12. If  $v$  is not continuous at  $x = s(t)$ , then the jump is called a *shock*, and the motion of this jump produces a *shock wave*. As shown in the next examples, the velocity of the shock is strongly dependent on the constitutive law.

*Example (Greenshields Law)* Using the linear law in (5.10), then the Rankine-Hugoniot condition (5.54) simplifies to the following:

$$\begin{aligned} s'(t) &= \frac{1}{\rho_R - \rho_L} \int_{\rho_L}^{\rho_R} v_M \left(1 - \frac{2\rho}{\rho_M}\right) d\rho \\ &= \frac{1}{2} (c_R + c_L). \end{aligned} \quad (5.55)$$

In other words, when using the Greenshields law, the shock moves at a speed determined by the average of the jump in the wave velocity across the shock. ■

*Example (Newell Law)* Using (5.18), then the Rankine-Hugoniot condition (5.54) is

$$\begin{aligned} s'(t) &= \frac{1}{\rho_R - \rho_L} [\rho_R v_M (1 - e_R) - \rho_L v_M (1 - e_L)] \\ &= v_M \left( 1 - \frac{\rho_R e_R - \rho_L e_L}{\rho_R - \rho_L} \right), \end{aligned}$$

where

$$\begin{aligned} e_L &= e^{-\lambda(1/\rho_L - 1/\rho_M)}, \\ e_R &= e^{-\lambda(1/\rho_R - 1/\rho_M)}. \end{aligned}$$

■

When we first started out studying traffic flow, we had only one variable with the dimension of velocity. Now, we have three variables with this dimension. They are:

1.  $v(x, t)$ . This is the velocity of the car at  $x$  at time  $t$  (using the averaging introduced earlier).
2.  $c(\rho)$ . This is the wave velocity, and it is defined in (5.35). It determines the slopes of the characteristic curves.
3.  $s'(t)$ . This is the velocity of the jumps in the solution, and it is given in (5.52).

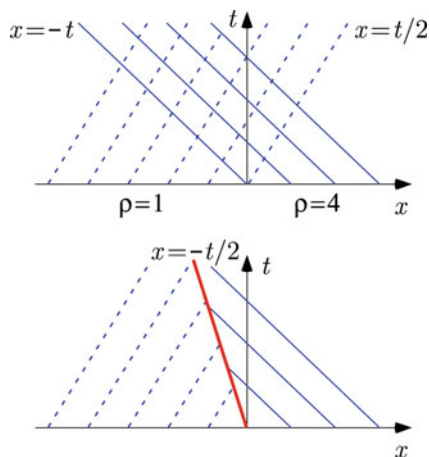
These velocities all play a critical role in the evolution of the solution and are distinct in the sense that, in most nonlinear problems, they are not simple multiples of each other. This is evident in the definitions of  $c$  and  $s'$ , as well as from the expressions derived in Exercise 5.12. What we conclude from this is that this interesting problem is rich enough that a single velocity is not enough to describe the solution.

### 5.6.4 Shock Waves

At a shock wave both the density and velocity are discontinuous. Calling the solution shown in Fig. 5.22 a shock wave gives the impression the cars are running into each other. They are not, and what happens when the shock passes over a car is that it immediately undergoes a jump in velocity. This is a bit unrealistic, and we will return to this point later.

Characteristics are used to determine when a shock wave is present in the solution. Namely, a shock appears when characteristics overlap, and the values on the characteristics are not equal. This situation is illustrated in the next example.

**Fig. 5.23** The traffic jam problem. The upper plot shows the characteristics associated with the initial condition. The lower plot shows the resulting shock location



*Example (Traffic Jam)* The problem is to solve

$$\frac{\partial \rho}{\partial t} + \left(1 - \frac{1}{2}\rho\right) \frac{\partial \rho}{\partial x} = 0, \quad (5.56)$$

with the initial condition

$$\rho(x, 0) = \begin{cases} 1 & \text{if } x < 0 \\ 4 & \text{if } x > 0. \end{cases} \quad (5.57)$$

This piecewise constant function gives rise to what is known as a *Riemann problem*. As in the earlier examples, the wave velocity used here comes from the Greenshields law, with  $v_M = 1$  and  $\rho_M = 4$ . Since  $c = 1 - \rho/2$ , then  $c_L = 1/2$  and  $c_R = -1$ , and the resulting characteristics are shown in upper graph in Fig. 5.23. We have a sector with overlapping characteristics. The left and right edges of the sector are determined by the characteristics  $x = -t$  and  $x = t/2$ , respectively. The solution in the sector will contain a shock as shown in the lower graph in Fig. 5.23. The solution on the left side of the shock is determined by the dashed characteristics, and the solution on the right is determined by the solid characteristics. As for the location of the shock, since we are using the Greenshields law, the velocity of the shock is determined using (5.55). So,  $s'(t) = -1/2$ , and since  $s(0) = 0$ , then  $s(t) = -t/2$ . Therefore, the solution of the problem is

$$\rho(x, t) = \begin{cases} 1 & \text{if } x < -t/2, \\ 4 & \text{if } -t/2 < x. \end{cases} \quad (5.58)$$



This example can be thought as a traffic jam problem. The reason is that the cars on the right are at the maximum density, and therefore do not move. The cars on the left, when they reach the jam, stop and the result is that the traffic jam spreads leftward along the negative  $x$ -axis. ■

The properties of the solution at a shock wave brings out one of the flaws in the traffic model. Specifically, as a shock passes over a car it immediately undergoes a jump in velocity. This is unrealistic and the reason it happens is that the model does not account for the momentum of the cars. Related to this is an assumption implicit in the constitutive law  $v = F(\rho)$ . For this to hold, the velocity must instantly adjust to the value of the density. This means that it is impossible to have the cars start from rest unless the density has a value where  $F(\rho) = 0$ . There are traffic models that do not have these complications, and one is the cellular automata model studied later in the chapter. Also, in the next chapter we will significantly extend the continuum model in such a way that momentum is a central component of the model.

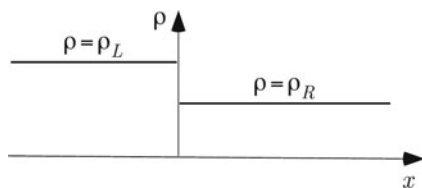
### 5.6.5 Expansion Fan

We will now consider what happens when the solution starts out with a jump but the slower cars are in the back. The initial condition is shown in Fig. 5.24, and it is given as

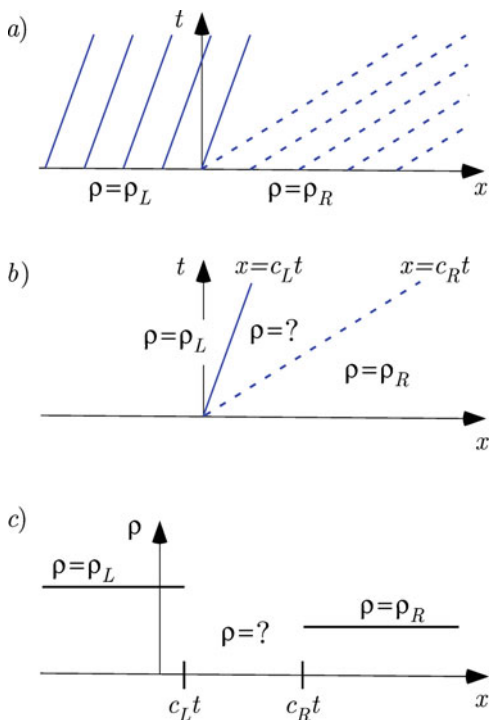
$$\rho(x, 0) = \begin{cases} \rho_L & \text{if } x < 0, \\ \rho_R & \text{if } 0 < x, \end{cases} \quad (5.59)$$

where  $0 < \rho_R < \rho_L$ . As mentioned earlier, this piecewise constant function gives rise to what is known as a *Riemann problem*. This particular example is interesting because the solution is not obvious. In fact, it is so unclear that it is possible to produce a plausible argument for at least three different solutions. Before describing what these are, we first state what we are certain of about the solution. This comes from the characteristics, and these are shown in Fig. 5.25a. As illustrated in Fig. 5.25b, c, we conclude that  $\rho = \rho_L$  for  $x < c_L t$  and  $\rho = \rho_R$  for  $x > c_R t$ . This leaves unresolved what the solution is for  $c_L t < x < c_R t$  because there are no characteristics in this region. It is what happens in this sector that produces the three possible solutions.

**Fig. 5.24** Initial density  $\rho(x, 0)$ , where the slow cars start out behind the faster cars (i.e.,  $\rho_L > \rho_R$ )



**Fig. 5.25** The solution obtained using the method of characteristics when the initial density is given in Fig. 5.24. As shown in (a) and (b), there are no characteristics in the sector  $c_L t < x < c_R t$ , and so the solution in that region is unclear



**Possibility 1.** All of the cars starting on the left, where  $x < 0$ , travel with velocity  $v_L$ , while those on the right have velocity  $v_R$ . Because  $v_L < v_R$ , then one might argue, based on physical grounds, that the sector in question is nothing more than the gap between the slow cars on the left and the fast cars on the right. In other words, for points in this sector the density is just zero and the apparent solution is

$$\rho(x, t) = \begin{cases} \rho_L & \text{if } x < v_L t, \\ 0 & \text{if } v_L t < x < v_R t, \\ \rho_R & \text{if } v_R t < x. \end{cases} \quad (5.60)$$

The first indication that there is something wrong with this expression is that the sector is determined by the velocity of the cars, and not the wave velocity. This is a problem because  $c(\rho_R) < v(\rho_R)$  and  $c(\rho_L) < v(\rho_L)$ , so the sector in (5.60) is different from the one shown in Fig. 5.25. In other words, the above expression contradicts what we certain of, which is the solution shown in Fig. 5.25c. Therefore, (5.60) is not the solution. It is also possible to use the solution of the modified red-light green-light problem to rule out this possibility, and how this is done is left as an exercise.

**Possibility 2.** As another attempt at finding out what happens in the sector one might argue that the solution of the linear traffic flow equation (5.19), using the

initial condition in (5.59), is a traveling wave with a single jump that moves with velocity  $a$ . Assuming the nonlinear equation also produces a single jump, then the apparent solution is a shock wave of the form

$$\rho(x, t) = \begin{cases} \rho_L & \text{if } x < s(t), \\ \rho_R & \text{if } s(t) < x. \end{cases} \quad (5.61)$$

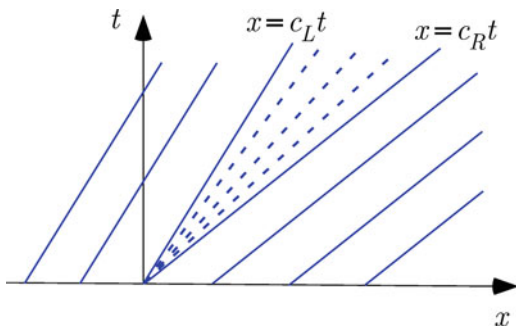
The function  $s(t)$  is determined from the Rankine-Hugoniot condition (5.54). Although it is not clear whether (5.61) is the solution, it has promise. For example, it is not hard to show that the line  $x = s(t)$  is between  $x = c_L t$  and  $x = c_R t$ . This means that (5.61) agrees with what we already know using characteristics, unlike what happened with (5.60). Moreover, in the special case of when  $c$  is constant, (5.61) reduces to the correct solution of the linear problem. These two observations are encouraging, but they do not guarantee that (5.61) is the solution of the Riemann problem we are studying.

**Possibility 3.** A third attempt at finding the solution makes use of the modified red-light green-light problem shown in Fig. 5.19. The solution of this modified problem should converge to the solution of our Riemann problem when  $\varepsilon \rightarrow 0$ . This, in effect, takes the dashed characteristics in Fig. 5.19 and pinches them together at the origin with the result shown in Fig. 5.26. The radial characteristics form what is known as an *expansion fan*, or rarefaction wave, and it connects the constant states on the left and right. The formula for the solution, which is obtained from (5.49), is, for  $t > 0$ ,

$$\rho(x, t) = \begin{cases} \rho_L & \text{if } x \leq c_L t, \\ \rho_L + (\rho_R - \rho_L) \frac{x - c_L t}{(c_R - c_L)t} & \text{if } c_L t < x < c_R t, \\ \rho_R & \text{if } c_R t \leq x. \end{cases} \quad (5.62)$$

The resulting solution looks much like the one in Fig. 5.19c in the sense that the expansion fan is responsible for a linear transition between the constant solutions on the left and right.

**Fig. 5.26** By letting  $\varepsilon \rightarrow 0$  the dashed characteristics in Fig. 5.19 form an expansion fan between  $x = c_L t$  and  $x = c_R t$



From the above discussion we have two contenders for the solution, namely (5.61) and (5.62). The fact that we have multiple possible solutions is because the nonlinear traffic flow problem is ill-posed, which in this case means that the problem is incomplete. What is required is an additional piece of information that will enable us to uniquely determine the solution. Moreover, it must be consistent with the physics of the problem. Exactly what this might mean depends on the application being considered.

The assumption used here is one of continuity. Namely, the jump appearing in the initial condition is almost impossible to produce physically, and in most experiments there is not a jump, but a small interval where the density changes in a rapid and continuous fashion from  $\rho_L$  to  $\rho_R$ . In this sense the initial condition containing a jump is simply a mathematical idealization of the true situation. Given that the solution with a continuous, but rapid, transition is known and given in Fig. 5.19, the condition we are searching for must be consistent with this result. In other words, the condition must be able to tell us that (5.62) is the solution to this problem.

There are various ways to write the needed condition, and we will use the one introduced by Lax (1973). The statement is that if the solution contains a jump, at  $x = s(t)$ , then the wave velocity behind the jump is larger than the wave velocity in front of it. This gives us the following assumption.

**Shock Admissibility Condition.** *For the solution of  $\partial_t \rho + c(\rho)\partial_x \rho = 0$  to have a shock at  $x = s(t)$ , it is required that*

$$c(\rho_R) < s' < c(\rho_L), \quad (5.63)$$

where  $\rho_L$  and  $\rho_R$  are the left and right values, respectively, of  $\rho$  across the shock.

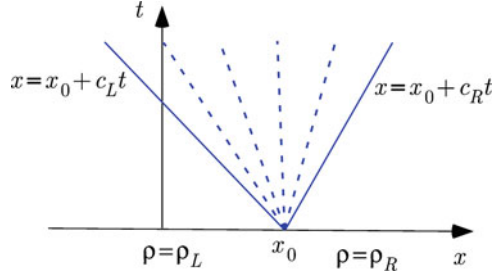
An immediate consequence of this condition is that the solution will only contain a shock if  $c_R < c_L$ . For our initial condition, given in (5.59), the assumption is that  $c_L < c_R$ . Therefore, a solution with a jump is not possible, and the solution in the region in question is an expansion fan. In other words, (5.62) is the solution of the stated Riemann problem. The proof of this statement can be found in Lax (1973).

There is a technical point to be made about the jump in the initial condition (5.59) as well as in the solution (5.61) containing a jump. As you will notice, the value at the jump is not specified. The reason is that there is not a consistent, or well-defined, value for the solution at such points. For example, with the expansion fan solution (5.62), by following the dashed characteristics into the origin one can get any value of  $\rho$  between  $\rho_L$  and  $\rho_R$ . This is similar to the situation we had in Sect. 4.4.2 when solving the diffusion equation with a jump in the initial condition.

### 5.6.5.1 General Formula for an Expansion Fan

The exact form of the expansion fan solution (5.62) relies on the specific formula for the wave velocity  $c(\rho)$  and a jump at  $x = 0$ . In general, a fan appears when there is a gap between characteristics as shown in Fig. 5.26. This occurs when  $f(x)$  has a

**Fig. 5.27** An expansion fan is generated at a point  $x_0$  where the initial function  $f(x)$  has a jump discontinuity, with  $c_R > c_L$



jump at a point  $x = x_0$ , with  $c_R > c_L$  (see Fig. 5.27). The equation for each of the dashed lines making up the fan has the form  $x = x_0 + c(\rho)t$ , where  $c(\rho)$  satisfies  $c_L < c < c_R$ . There are a couple of methods that can be used to prove this, other than taking a limit as we did earlier, and one is explored in Exercise 5.30.

To determine the density at a point  $(x, t)$  in the fan, it is necessary to find the value of  $\rho$  so that  $x = x_0 + c(\rho)t$ . In other words, it is required to solve the equation

$$c(\rho) = \frac{x - x_0}{t} \quad (5.64)$$

for  $\rho$ . How easy, or difficult, it is to find  $\rho$  depends on the specific form of  $c$ . Also, in formulating the nonlinear traffic flow equation in Sect. 5.6, we stated that it is assumed that  $c(\rho)$  is continuous and monotonic. This is one of the places where we need these assumptions because they guarantee that (5.64) has a unique solution. As a final comment, in the case of when a gap between the characteristics occurs for  $t = t_0$ , then the denominator in (5.64) is replaced with  $t - t_0$ .

*Example (Red Light–Green Light)* The problem is to solve

$$\frac{\partial \rho}{\partial t} + (2 - \rho) \frac{\partial \rho}{\partial x} = 0, \quad (5.65)$$

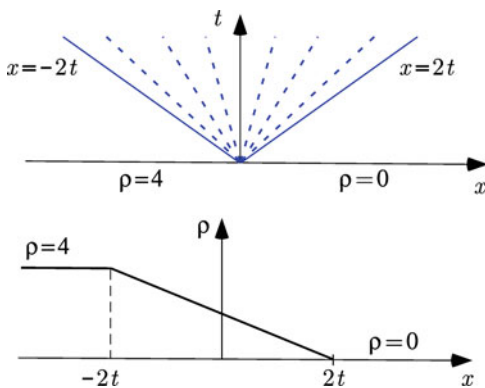
with the initial condition

$$\rho(x, 0) = \begin{cases} 4 & \text{if } x < 0 \\ 0 & \text{if } x > 0. \end{cases} \quad (5.66)$$

The wave velocity  $c = 2 - \rho$  comes from the Greenshields law, with  $v_M = 2$  and  $\rho_M = 4$ .

*Initial Jump at  $x = 0$ :* For the jump in the initial condition at  $x = 0$ ,  $\rho_L = 4$ ,  $c_L = -2$ ,  $\rho_R = 0$ , and  $c_R = 2$ . In this case, since  $c_L < c_R$ , an expansion fan is generated by this jump. The left and right edges of the fan are determined from the characteristics  $x = x_0 + c_L t = -2t$  and  $x = x_0 + c_R t = 2t$ , respectively.

**Fig. 5.28** The upper plot shows the solution on the left and right, and the characteristics for the expansion fan. The lower plot shows the solution after the light turns green



To find the solution in the fan, from (5.64), it is required that  $2 - \rho = \frac{x}{t}$ . Solving this yields

$$\rho = 2 - \frac{x}{t}.$$

Therefore, the solution of the traffic flow equation is, for  $t > 0$ ,

$$\rho(x, t) = \begin{cases} 4 & \text{if } x \leq -2t, \\ 2 - \frac{x}{t} & \text{if } -2t < x < 2t, \\ 0 & \text{if } 2t \leq x. \end{cases} \quad (5.67)$$

The solution is shown in Fig. 5.28, along with the associated characteristic curves. This shows that when the light turns green the cars move to the right, with the front moving at the maximum allowable velocity  $v_M$ . ■

### 5.6.5.2 Parting Comments

What is interesting about the admissibility condition is how it relies on the physical problem to determine which solution to use. There are more subtle complications related to picking solutions for the traffic flow problem, and one is discussed in Exercise 5.33.

After reading the above paragraphs one might decide that the best thing to do is avoid using an initial condition with a jump. However, as indicated in Fig. 5.21, and as shown in the next section, this nonlinear equation can take a continuous initial condition and cause it to form jumps. So, even if we do not feed it jumps at the beginning it can easily grow its own and this means there is no avoiding having to consider an admissibility condition.

Because the admissibility condition that one uses has the potential to generate some unusual solutions, it is worth discussing this a bit more. The approach used here to help determine which solution to use involved modifying the initial condition (and not the differential equation). In contrast, in more mathematical studies the underlying physical assumptions are often modified. For example, one approach is to assume that diffusion is also present, with the result that the equation becomes  $\partial_t \rho + \partial_x(\rho v) = \varepsilon \partial_x^2 \rho$ . What is of interest in this case is what happens when  $\varepsilon \rightarrow 0$ . This approach has its origins in gas dynamics, where the diffusion term can be associated with the viscous forces in the gas. Such an interpretation is not really possible for traffic flow, or other applications not directly connected to fluid flow. This has certainly not stopped research on this subject, and those interested in investigating this topic should consult Gasser (2003) and Knowles (2008). There has also been interest in what happens to the admissibility condition when not just diffusion is used, but dispersion or other physical effects are included. This can give rise to something called an undercompressive shock, and more about this can be found in El et al. (2017).

### 5.6.6 Summary

The results we have derived for the traffic flow problem are scattered through the preceding pages, and it is worth collecting them together. The problem consists of the first-order partial differential equation

$$\frac{\partial \rho}{\partial t} + c(\rho) \frac{\partial \rho}{\partial x} = 0, \quad (5.68)$$

with the initial condition

$$\rho(x, 0) = f(x). \quad (5.69)$$

Assuming  $c(\rho)$  is a smooth function, with  $c'(\rho) \neq 0$ , for  $0 \leq \rho \leq \rho_M$ , then the solution is constructed using the following information.

- (a) The solution is constant along the characteristic curves  $x = x_0 + c_0 t$  (see Fig. 5.17).
- (b) Characteristics Overlap. In a sector containing overlapping characteristics the solution contains a shock wave at  $x = s(t)$ . The velocity of this wave is

$$s'(t) = \frac{1}{\rho_R - \rho_L} \int_{\rho_L}^{\rho_R} c(\rho) d\rho. \quad (5.70)$$

On either side of the shock, the respective characteristics determine the solution (as illustrated in Fig. 5.23). As an example, if  $f(x)$  is piecewise constant with a

jump discontinuity at  $x_0$ , with  $c_R < c_L$ , then the solution starts out with a shock wave of the form  $x = x_0 + s'_0 t$ , where  $s'_0$  is determined from (5.70).

- (c) **Characteristics Separate.** In a sector with no characteristics, the solution is an expansion fan. An example is shown in Fig. 5.27, where  $f(x)$  has a jump discontinuity at  $x_0$ , with  $c_R > c_L$ . In this case, the left and right edges of the fan are determined from the characteristics  $x = x_0 + c_L t$  and  $x_0 + c_R t$ , respectively. Moreover, for points  $(x, t)$  located in the fan, the solution is found by solving  $c(\rho) = (x - x_0)/t$  for  $\rho$ .

The above conclusions are general in the sense that they apply to traffic flow, where  $c' < 0$ , but also to the case of when  $c' > 0$ . The latter occurs, for example, for gas flow, which is the subject of Exercise 5.26.

### 5.6.7 Additional Examples

In what follows, examples are considered which combine the various solutions we have considered earlier. In each case the Greenshields law is used, which means that the wave velocity is  $c = v_M(1 - 2\rho/\rho_M)$ . Also, the shock velocity is  $s' = \frac{1}{2}(c_L + c_R)$ .

*Example (Shock and Shock)* Suppose the initial condition is

$$\rho(x, 0) = \begin{cases} 1 & \text{if } x < 0 \\ 2 & \text{if } 0 < x < 2 \\ 4 & \text{if } 2 < x. \end{cases} \quad (5.71)$$

Also, taking  $v_M = 1$  and  $\rho_M = 12$ , then  $c = 1 - \rho/6$ . The shock solutions derived below are shown in the upper graph in Fig. 5.29.

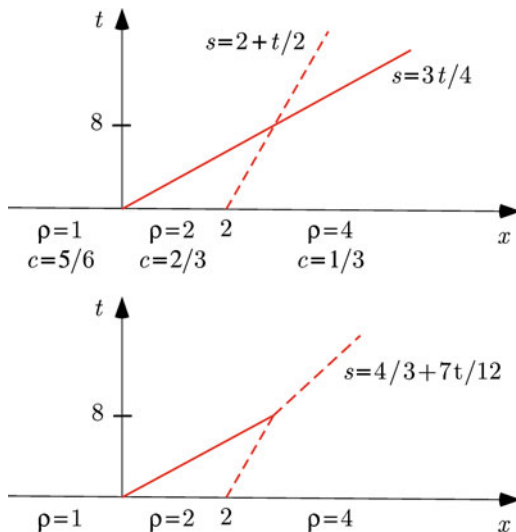
*Initial Jump at  $x = 0$ :* For the jump in the initial condition at  $x = 0$ , we have  $\rho_L = 1$ ,  $c_L = 5/6$ ,  $\rho_R = 2$ , and  $c_R = 2/3$ . Because  $c_L > c_R$ , then there is a shock generated by this jump, and it has velocity  $s' = 3/4$ . Since  $s(0) = 0$ , we get that  $s(t) = 3t/4$ .

*Initial Jump at  $x = 2$ :* For the jump in the initial condition at  $x = 2$ , we have  $\rho_L = 2$ ,  $c_L = 2/3$ ,  $\rho_R = 4$ , and  $c_R = 1/3$ . Because  $c_L > c_R$ , then there is a shock generated by this jump, and it has velocity  $s' = 1/2$ . Since  $s(0) = 2$ , we get that  $s(t) = 2 + t/2$ .

As can be seen in Fig. 5.29, the shocks intersect, and this occurs at  $(x, t) = (6, 8)$ . The sector formed by the two shocks, for  $t > 8$ , contains overlapping characteristics (these are not shown in the figure but the situation is similar to what is shown in Fig. 5.21). Consequently, a new shock forms at the intersection point. Since  $\rho_L = 1$ ,  $c_L = 5/6$ ,  $\rho_R = 4$ , and  $c_R = 1/3$ , then  $s' = 7/12$ . Given that  $s(8) = 6$ , it follows that the newly formed shock is given as  $s(t) = 4/3 + 7t/12$ .



**Fig. 5.29** Upper: Shocks generated by initial condition (5.74). Lower: Shock generated when the two original shocks meet



Summarizing the above results, for  $0 \leq t < 8$ , the solution is

$$\rho(x, t) = \begin{cases} 1 & \text{if } x < 3t/4 \\ 2 & \text{if } 3t/4 < x < 2 + t/2 \\ 4 & \text{if } 2 + t/2 < x. \end{cases} \quad (5.72)$$

For  $8 \leq t$  the solution is

$$\rho(x, t) = \begin{cases} 1 & \text{if } x < 2 + t/3 \\ 4 & \text{if } 2 + t/3 < x. \end{cases} \quad (5.73)$$

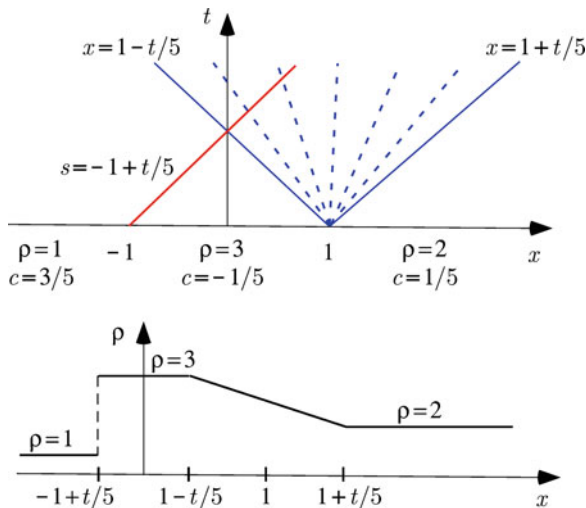
The solution we have derived is not a surprise. According to the initial condition (5.74), the fastest cars are in the back, and the slowest cars are in the front. So, two shocks form immediately between the three groups (this is the solution given in (5.72)). Eventually the fastest group in the back overtakes all of the middle group, at which point the resulting jump in the solution is between the fastest and slowest groups (this is the solution given in (5.73)). ■

*Example (Shock and Fan)* Suppose the initial condition is

$$\rho(x, 0) = \begin{cases} 1 & \text{if } x < -1 \\ 3 & \text{if } -1 < x < 1 \\ 2 & \text{if } 1 < x. \end{cases} \quad (5.74)$$

Also, taking  $v_M = 1$  and  $\rho_M = 5$ , then  $c = 1 - 2\rho/5$ .

**Fig. 5.30** Upper: Shock and expansion fan generated by initial condition (5.74). Lower: Solution before the shock hits the fan at  $t = 5$



*Initial Jump at  $x = -1$ :* For the jump in the initial condition at  $x = -1$ ,  $\rho_L = 1$ ,  $c_L = 3/5$ ,  $\rho_R = 3$ , and  $c_R = -1/5$ . Because  $c_L > c_R$ , then there is a shock generated by this jump, and it has velocity  $s' = 1/5$ . Since  $s(0) = -1$ , we get that  $s(t) = -1 + t/5$ . This is the red line shown in Fig. 5.30.

*Initial Jump at  $x = 1$ :* For the jump in the initial condition at  $x = 1$ ,  $\rho_L = 3$ ,  $c_L = -1/5$ ,  $\rho_R = 2$ , and  $c_R = 1/5$ . In this case, since  $c_L < c_R$ , an expansion fan is generated by this jump. The left and right edges of the fan are determined from the characteristics  $x = 1 - t/5$  and  $x = 1 + t/5$ , respectively. Also, from (5.64), the solution in the fan is, for  $t > 0$ ,

$$\rho = \frac{5}{2} \left( 1 - \frac{x-1}{t} \right).$$

An illustration of the solution, so far, is given in Fig. 5.30. It is seen that the shock wave eventually hits the fan, and this occurs when  $t = 5$  (and  $x = 0$ ). So, for  $0 < t < 5$ , the solution is

$$\rho(x, t) = \begin{cases} 1 & \text{if } x < -1 + t/5 \\ 3 & \text{if } -1 + t/5 < x < 1 - t/5 \\ \frac{5}{2} \left( 1 - \frac{x-1}{t} \right) & \text{if } 1 - t/5 \leq x \leq 1 + t/5 \\ 2 & \text{if } 1 + t/5 < x. \end{cases}$$

As shown in Fig. 5.30, for  $t > 5$ , the shock overlaps with the characteristics in the fan. Consequently, for  $t > 5$  a shock continues but its velocity is affected by the changing value of  $c_R$  as it works its way across the fan. It eventually reaches the  $\rho = 2$  solution, after which its velocity is  $s' = 2/5$ . The formulas for the shock as it passes through the fan, as well as afterwards, are derived in Exercise 5.16.

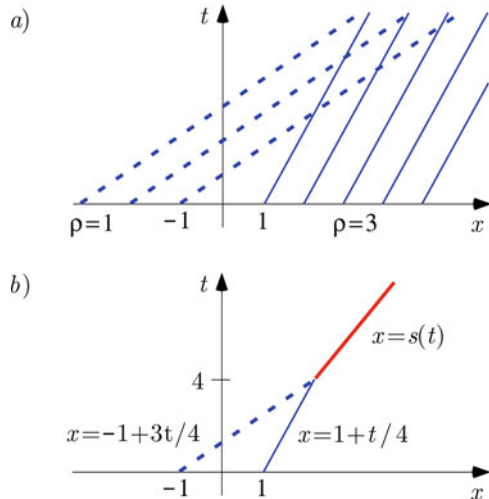
The solution we have found is easy to understand physically. Namely, we have placed the fastest cars in the back and the slowest cars in the middle. So, a shock immediately forms at the interface between these two groups (this is the jump in  $\rho$  shown in the lower graph in Fig. 5.30). The cars in the front ( $\rho = 2$ ) are also faster than those in the middle, and so these two groups begin to separate. The transition region between them is the expansion fan. Eventually the faster cars on the left ( $\rho = 1$ ) overtake all of the middle group. When this happens, the jump in  $\rho$  occurs where the  $\rho = 1$  cars are overtaking those in the expansion fan. From this point on the shock (jump) works itself across the expansion fan, and it eventually catches up to the  $\rho = 2$  cars. Once this happens the resulting solution is a single shock wave, separating the  $\rho = 1$  and  $\rho = 2$  groups, propagating to the right. ■

*Example (No Initial Jumps)* As a third example suppose the density does not begin with a jump, but is continuous and is given as

$$\rho(x, 0) = \begin{cases} 1 & \text{if } x \leq -1 \\ 2 + x & \text{if } -1 < x < 1 \\ 3 & \text{if } 1 \leq x. \end{cases} \quad (5.75)$$

Also, taking  $v_M = 1$  and  $\rho_M = 8$ , then  $c = 1 - \rho/4$ . The characteristics that are produced by the two constant values on the left and right are shown in Fig. 5.31a. In the region covered by the dashed lines the solution is  $\rho = 1$ , while in the region covered by the solid lines the solution is  $\rho = 3$ . The exception to this statement is where the dashed and solid lines overlap, which indicates the formation of a shock.

**Fig. 5.31** Overlapping characteristics are shown in (a), which indicates the existence of a shock wave in this region. The position of the shock is shown in (b), along with the two characteristics that intersect to initiate the formation of the shock at  $t = 4$



To determine this, note that the dashed characteristic starting at  $x = -1$  is  $x = -1 + 3t/4$ , and the solid characteristic beginning at  $x = 1$  is  $x = 1 + t/4$ . These lines intersect at  $(x, t) = (2, 4)$ . For the resulting shock,  $s' = \frac{1}{2}(c_L + c_R)$ , where  $s(4) = 2$ . Since  $c_L = 3/4$  and  $c_R = 1/4$ , it follows that

$$s(t) = \frac{1}{2}t, \text{ for } 4 \leq t. \quad (5.76)$$

It remains to determine the solution in the triangular region shown in Fig. 5.31b, which is bounded by the characteristics  $x = -1 + 3t/4$  and  $x = 1 + t/4$ . Even though we have a formula for the solution, given in (5.48), we will derive it using the particular values for this example. So, given a point  $(x, t)$  in this region, we need to find the characteristic that passes through it. This means we need to find the value of  $x_0$  so that

$$x = x_0 + c_0 t.$$

Since  $c_0 = 1 - \rho_0/4 = 1 - (2 + x_0)/4 = 1/2 - x_0/4$ , then we must solve  $x = x_0 + (1/2 - x_0/4)t$  for  $x_0$ . This yields the solution

$$\rho(x, t) = \rho(x_0, 0) = 4 \frac{2 + x - t}{4 - t}.$$

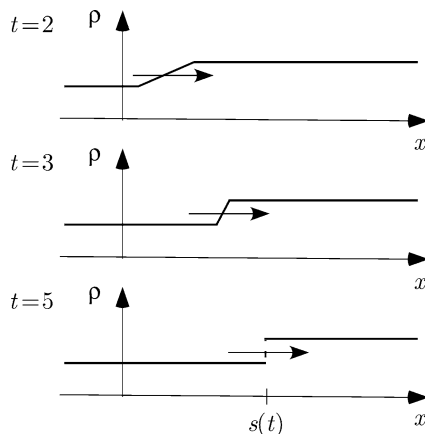
We have therefore found that the solution for  $0 \leq t < 4$  is

$$\rho(x, t) = \begin{cases} 1 & \text{if } x \leq 1 + 3t/4, \\ 4 \frac{2+x-t}{4-t} & \text{if } 1 + 3t/4 < x < 1 + t/4, \\ 3 & \text{if } 1 + t/4 \leq x, \end{cases} \quad (5.77)$$

and for  $4 \leq t$  the solution is

$$\rho(x, t) = \begin{cases} 1 & \text{if } x < \frac{1}{2}t, \\ 3 & \text{if } \frac{1}{2}t < x. \end{cases} \quad (5.78)$$

This solution is shown in Fig. 5.32 at three time values. At  $t = 2$  the solution consists of the two constant densities that are connected by a linear function. Because the cars on the left are faster than those on the right, at the later time  $t = 3$  the linear connection between the two densities has been reduced considerably. The effect of this transition region shrinking is to steepen the wave as it moves. The faster cars eventually catch the slower ones in front, at which point a shock forms. This is the solution seen at time  $t = 5$ . ■



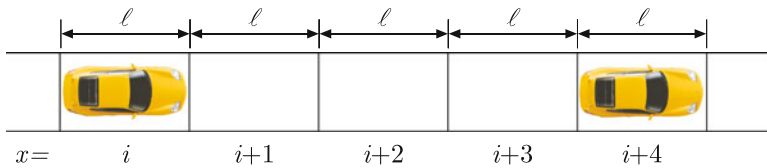
**Fig. 5.32** The solution of the traffic flow problem at the times shown in Fig. 5.19b. The width of the linear transition region between the left and right groups decreases with time until the left group catches the right group, giving rise to a shock wave

## 5.7 Cellular Automata Modeling

The viewpoint of the continuum model derived in Sect. 5.3 is that the motion of the individual cars can be approximated using an averaging process, giving rise to the density and flux functions. It is interesting to explore how to retain the individuality of the cars, and one approach incorporates ideas from cellular automata. The first step in constructing the model is to divide the road into equal segments, each with length  $\ell$  as shown in Fig. 5.33. Conventionally, this distance is taken to be the length of an average car, or vehicle, on the road. Time is also divided into equal segments, producing a time step  $\Delta t$ . The objective of the model is, given the positions of the cars at time  $t_{old}$ , to determine their positions at  $t_{new} = t_{old} + \Delta t$ . With this in mind we introduce an integer variable  $m$  that equals the number of road segments the car moves in a time step. For example,  $m = 1$  means the car moves one segment,  $m = 2$  means it moves two segments, etc. It is assumed that there is a maximum number of segments  $M$  that a car is allowed to move in a time step. This is equivalent to assuming there is a maximum velocity  $v_M$  on the highway. Given that a car's velocity in this formulation is  $v = m\ell/\Delta t$ , then  $v_M = M\ell/\Delta t$ .

*Example* For a typical passenger car,  $\ell = 16$  ft (4.9 m). Taking  $\Delta t = 1$  s, then  $m = 1$  corresponds to a velocity of 10.9 mph (17.5 kph), while  $m = 6$  corresponds to a velocity of 65.4 mph (105.2 kph). ■

In the model, each car has three integers associated with it, and they are  $(x, m, g)$ . Here  $x$  is the position of the car and its value is determined by the road segment currently occupied by the car. The integer  $m$  was defined earlier, and  $g$  is called the gap and it is the number of spaces between the car and the one in front of it. For example, for the car on the left in Fig. 5.33,  $x = i$  and  $g = 3$ , while for the car on the right  $x = i + 4$ .



**Fig. 5.33** In traffic cellular automaton models the roadway is divided into equal segments, and the segments are numbered. For the car on the left,  $x = i$  and its gap, which is the number of empty segments in front of it, is  $g = 3$

The basic idea in the model is that at time  $t_{old}$  we know the values of  $(x_{old}, m_{old}, g_{old})$  for each car, and what the model does is to determine their values  $(x_{new}, m_{new}, g_{new})$  at time  $t_{new} = t_{old} + \Delta t$ . This is done by applying the following four rules to each car on the road:

1. *Speedup.*  
If  $m_{old} \neq M$ , then  $m_{new} = m_{old} + 1$ .
2. *Do Not Overrun.*  
If  $m_{new} > g_{old}$ , then  $m_{new} = g_{old}$ .
3. *Randomization.*  
If  $m_{new} \neq 0$ , then, with probability  $p$ , take  $m_{new} = m_{new} - 1$ .
4. *Move the Car.*  
Take  $x_{new} = x_{old} + m_{new}$ .

These four steps constitute what is known as the Nagel-Schreckenberg, or NaSch, model (Nagel and Schreckenberg, 1992). It is an example of what is known as a stochastic cellular automata model for traffic flow.

The first three steps of the above model contain assumptions, and potential modifications, that need to be discussed.

*Speedup* It is assumed that a driver will attempt to drive at the maximum allowed velocity. Assuming the car is not moving at the maximum velocity, then in this step the number of segments a car moves is increased by one. It is certainly possible to consider what happens if there are larger accelerations, and increase the movement by two or more segments, but this will not be investigated here.

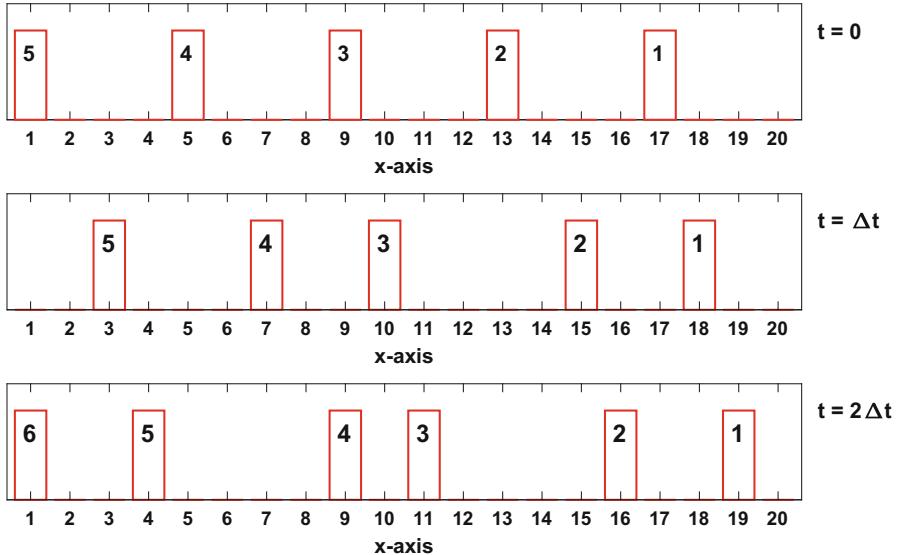
*Do Not Overrun* The idea here is that if there are, say, three empty spaces in front of the car, then it cannot move any more than three spaces in the time step. One can argue that the car in front will likely move in the time step and therefore there will be more than three available spaces. This is correct but it is not accounted for in the model.

*Randomization* Numerous reasons have been given to justify this assumption, and this includes the statements that it mimics delayed acceleration or that it accounts for an overreaction in braking. In this vein,  $p$  is referred to as the dawdling probability.

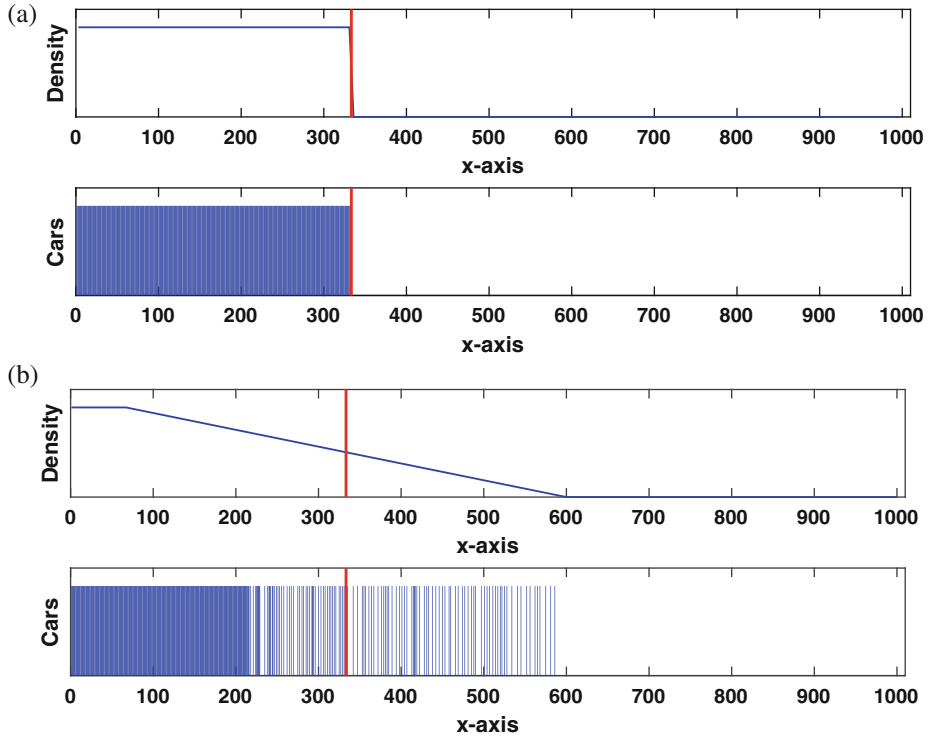
Given the recursive nature of how the four rules are used, it is difficult to determine exactly what happens using analytical methods. The approach, therefore, is to use computer simulations and this brings us to the next example.

*Example* Suppose the cars start out uniformly distributed along the highway with a gap  $g = 3$ . This is shown in Fig. 5.34. The lower two graphs are the positions of the cars at the first two time steps, assuming  $p = 2/3$ ,  $m = 1$ , and  $M = 2$ . At  $t = \Delta t$ , the speedup step moves each car two spaces, which happens for cars 5, 4, and 2. However, the randomization step affects cars 3 and 1, and they only move one space. At  $t = 2\Delta t$ , car 4 is the only one that moves the maximum number of two steps, the others being affected by the randomization step. What is new at this time step is the appearance of a sixth car on the left. This is not from the model but, rather, something that is included in the computer code. Specifically, if the first space is empty, then the computer adds in a car at this location with a probability equal to the original uniform distribution, namely with probability  $1/(g + 1)$ . ■

It is possible to take the computer code that produced Fig. 5.34 and calculate car positions using a large number of road segments and time steps. Although this generates interesting pictures, little is learned in the process. It is better to study specific situations and compare the results with what is expected on a real roadway. For this we turn to the red light–green light and the green light–red light examples introduced earlier for the continuum model.



**Fig. 5.34** Five cars are placed uniformly along a roadway at  $t = 0$ . Their positions calculated using the NaSch model are shown at  $t = \Delta t$  and at  $t = 2\Delta t$



**Fig. 5.35** Solution of the red light–green light problem. Show are (a) the density and positions of the cars at  $t = 0$ , and (b) the density and positions after several time steps. The density is computed using (5.79) and the car positions are determined using the NaSch model

*Example (Red Light–Green Light)* For this the road is divided into 1000 segments, and the stoplight is located at  $x = 333$ . It is assumed the cars are bumper-to-bumper to the left of the light. This is shown in the lower plot of Fig. 5.35a, where the solid blue block on the left are the 333 cars waiting for the light to turn green. Shown in the upper plot is the corresponding density, using the continuum model.

The solution at  $t = N\Delta t$ , where  $N = 133$  and  $\Delta t = 1$ , for both models is shown in Fig. 5.35b. In the calculation  $M = 2$ ,  $p = 1/10$ ,  $\ell = 1$ ,  $\rho_L = \rho_M = 1$ ,  $\rho_R = 0$ , and  $v_M = 2$ . For the NaSch model the cars on the right have pulled ahead while those in the block on the left are waiting for room to open up so they can move. For the continuum model we have an expansion fan. From (5.62), in the case of when the light is located at  $\bar{x} = 333$ , the solution is

$$\rho(x, t) = \begin{cases} 1 & \text{if } x \leq \bar{x} - 2t, \\ \frac{2t + \bar{x} - x}{4t} & \text{if } \bar{x} - 2t < x < \bar{x} + 2t, \\ 0 & \text{if } \bar{x} + 2t \leq x. \end{cases} \quad (5.79)$$



The linear transition between  $\rho = 1$  and  $\rho = 0$  is seen in the upper plot of Fig. 5.35b.

We are now able to ask the big question, namely, how do the two models compare? To investigate this note that for the NaSch model the car in the front, after  $N$  time steps, can move no farther than  $MN$  spaces. The actual number is smaller, depending on how many time steps it takes to accelerate to velocity  $M$  and the always present reduction of the velocity due to the randomization step. If  $p$  is close to zero, then the first car moves approximately  $N_M = M[N - \frac{1}{2}(M - 1)]$  spaces, where the  $\frac{1}{2}M(M - 1)$  term is due to the number of steps it takes the car to reach speed  $M$ . In Fig. 5.35b this car is located at about  $x = 590$ . The leftmost car that is able to move, for  $p$  close to zero, is located approximately  $N$  spaces to the left of the light, which in Fig. 5.35b is near  $x = 220$ . At the other extreme, the closer  $p$  gets to one the closer the number of spaces on either the left or right gets to zero. In this discussion we will assume a linear approximation in  $p$ . Therefore, the approximate spatial interval involving the cars that are in motion after  $N$  time steps is

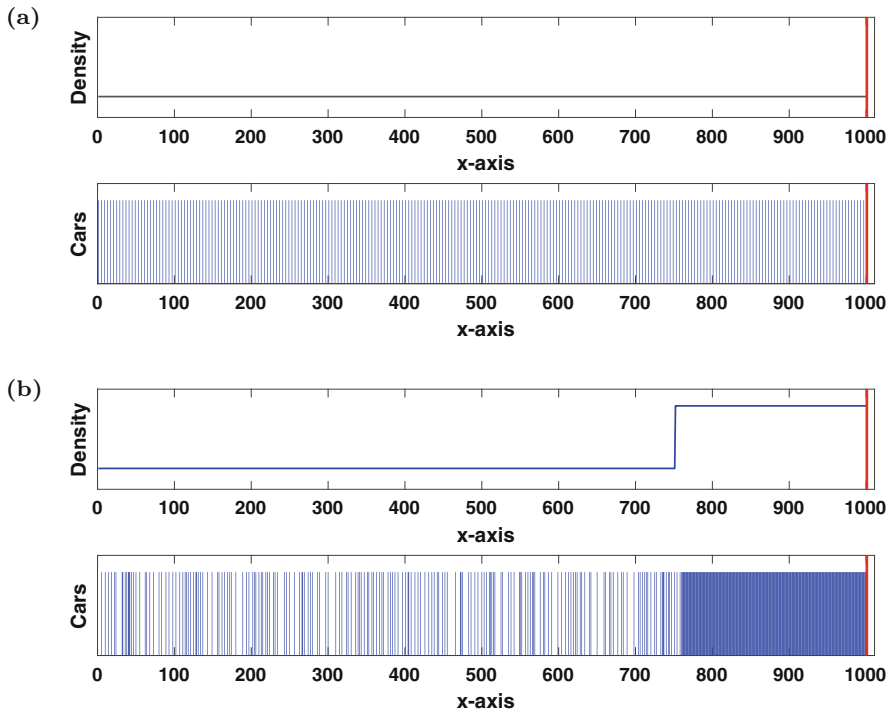
$$\bar{x} - N(1 - p) \leq x \leq \bar{x} + N_M(1 - p). \quad (5.80)$$

In comparison, for the continuum model the interval is

$$\bar{x} - 2N \leq x \leq \bar{x} + 2N. \quad (5.81)$$

This result brings out one clear difference between the two models. Specifically, the continuum model predicts the interval is symmetric about the light's position, while for the NaSch model the interval is nonsymmetric. This is seen in Fig. 5.35b. The symmetry in the continuum model is due to the Greenshields law which, for this example, gives  $c_L = -c_R$ . This means the front and back of the fan are moving in opposite directions with the same speed, hence the symmetry. For more physically realistic constitutive laws the speed is generally not symmetric, in which case we would obtain a nonsymmetric interval. Can we find a constitutive law that produces the same result as the NaSch model? Well, this rather interesting question will be left for you to think about. ■

*Example (Green Light–Red Light)* For this example the road is again divided into 1000 segments, with a stoplight located at  $x = 1000$ . It is assumed that the cars are uniformly spaced with three spaces between them, and each starts out at the maximum velocity  $v_M = 2$ . The randomization probability in this example is  $p = \frac{1}{4}$ , and the other parameters are the same as in the previous example. The initial positions are shown in Fig. 5.36a. Also shown in this figure is the corresponding density  $\rho = 1/4$  for the continuum model. The solution after a few time steps, for both the NaSch and the continuum models, is shown in Fig. 5.36b. Both are behaving as expected. In the NaSch model the cars coming in from the left stop when they arrive at the traffic jam, which in Fig. 5.36b is located near approximately  $x = 770$ . For the continuum model we have a shock wave that moves leftward.



**Fig. 5.36** Solution of the green light–red light problem. Shown are (a) the density and positions of the cars at  $t = 0$ , and (b) the density and positions after several time steps. The density is computed using (5.58) and the car positions are determined using the NaSch model

Because  $\rho_L = 1/4$  and  $\rho_R = 1$ , then, after  $N$  time steps, the shock is located at  $s = \bar{x} - N/2$ , where  $\bar{x}$  is the location of the stoplight. In Fig. 5.36b, where  $N = 450$ , we have that  $s = 780$ .

One of the more obvious differences in the two models, when looking at Fig. 5.36b, is the lack of uniformity in the density to the left of the traffic jam in the NaSch description. This is not unexpected and is due to the randomization step. The second difference is the location of the traffic jam. It is difficult to predict where the jam is located using the NaSch model because there is no simple formula as there is in the continuum case. What is interesting in Fig. 5.36b is that it appears that in the NaSch model the jam affects the motion of the cars before they reach the jam. Specifically, there is an increase in the density as the cars approach the jam. This can be explained by the fact that any time a car slows down this information is sent backwards along the road (Step 2). Therefore, when a car slows down as it arrives at the jam this affects the cars that follow. This is not in the continuum model and consequently represents a fundamental difference between the two descriptions. ■

## Exercises

### Sections 5.2–5.4

**5.1** This problem considers various consequences of the traffic flow equation.

- (a) Show that given any two points  $a$  and  $b$  on the  $x$ -axis, with  $a < b$ ,

$$\frac{d}{dt} \int_a^b \rho(x, t) dx = J(\rho_a) - J(\rho_b),$$

where  $\rho_a = \rho(a, t)$  and  $\rho_b = \rho(b, t)$ . Interpret the above equation in physical terms.

- (b) Assuming  $\rho(x, 0) = f(x)$ , show that

$$\rho(x, t) = f(x) - \frac{\partial}{\partial x} \int_0^t J(x, \tau) d\tau.$$

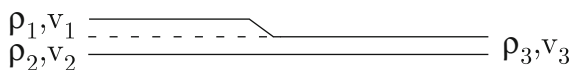
**5.2** Consider the situation of when two lanes of traffic merge down to one lane, as shown in Fig. 5.37. Assume a steady flow, so the density and velocity do not depend on time, and assume all variables are nonnegative.

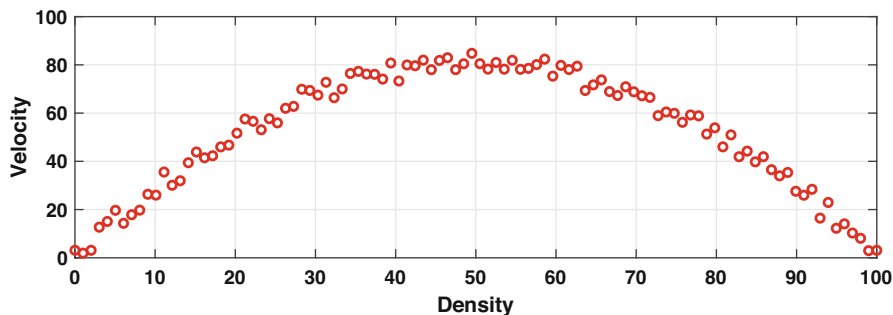
- (a) Using the result of Exercise 5.1a, find an equation that relates the values on the right with those on the left.  
 (b) What does the equation in part (a) reduce to if the Greenshields law is used?  
 (c) Suppose  $\rho_2 = \rho_1$ ,  $v_2 = v_1$ , and the Greenshields law is used. Find  $\rho_3$  in terms of  $\rho_1$ . Your solution should give  $\rho_3 = 0$  if  $\rho_1 = 0$ .  
 (d) Using your result from part (c), describe what happens to the flow of cars on the right as  $\rho_1$  is increased, starting from  $\rho_1 = 0$ . Make sure to explain what happens as  $\rho_1$  nears  $\frac{1}{4}(2 - \sqrt{2})\rho_M$ .

**5.3** There are various recommendations concerning safe following distances for cars. Below are a few of the more commonly cited rules. Find the resulting constitutive law relating density and velocity if you assume the cars are uniformly spaced according to the given rule. The function  $F(\rho)$  must be continuous, and if you need to make additional assumptions to derive the requested constitutive law make sure to state what they are.

- (a) The National Safety Council recommends the three-second rule. This means that you allow at least 3 s between you and the vehicle in front of you.  
 (b) In the early days of motoring, it was recommended that you keep one car length back (about 20 ft) for each 10 mph of speed.

**Fig. 5.37** Configuration of roadway used for Problem 5.2





**Fig. 5.38** Data used for Problem 5.4

- (c) According to an insurance company, you should allow at least 4 s between you and the vehicle in front of you, but if traveling more than 50 mph that this time interval should be at least 6 s.
- (d) According to a motoring society, the minimum safe distance to the car in front is made up of the sum of two terms. One accounts for the distance traveled due to reaction time, which is usually assumed to be 1.5 s. The second term is calculated assuming a constant deceleration, and it accounts for the distance the car will travel after the brakes are applied.

**5.4** In the traffic flow problem suppose the velocity of cars, as a function of the density, is measured on a highway and the data shown in Fig. 5.38 are obtained.

- (a) Formulate a constitutive law for  $v$  as a function of  $\rho$  based on these data. Provide an explanation of how you reach your conclusion.
- (b) In the traffic model it is assumed that  $c'(\rho) \neq 0$  for  $0 \leq \rho \leq \rho_M$ . Does your constitutive law satisfy this condition?

**5.5** This problem explores some of the consequences of the Greenshields model as identified in a typical traffic engineering manual.

- (a) Sketch the flux as a function of density. At what density is the flux a maximum?
- (b) The constant  $\rho_M$  is called the jam density,  $v_M$  is called the free-flow velocity, and  $\frac{1}{4}v_M\rho_M$  is the capacity. Explain why they are given these names.
- (c) The headway is defined as the distance between a common point on adjacent cars (e.g., the distance between the front bumpers). The time headway is the time it takes the car to transverse this distance. How are these related to the flux or velocity?
- (d) In the example of Sect. 5.2 for uniform cars the maximum merge density  $\rho_{merge}$  was calculated. Use this and the data in Fig. 5.6 to find an approximate value for the maximum merge velocity  $v_{merge}$ , which is the velocity corresponding to the maximum merge density.

**5.6** A variable related to density is the volume fraction  $\phi(x, t)$ , which is used to determine how much of the highway is taken up by cars (versus empty road). In reference to Fig. 5.4,

$$\phi(x_0, t_0) \approx \frac{\text{total length of cars from } x_0 - \Delta x \text{ to } x_0 + \Delta x \text{ at } t = t_0}{2\Delta x}.$$

The value of  $\phi(x_0, t_0)$  is the limit of the right-hand side as  $\Delta x \rightarrow 0$ .

- (a) For evenly spaced cars as in Example 5.2, show that  $\phi(x, t) = \ell/(\ell + d)$ , and therefore  $\phi = \ell\rho$ .
- (b) If the cars are not necessarily evenly spaced but still are all of length  $\ell$  show that it is still true that  $\phi = \ell\rho$ .
- (c) Assuming  $\phi = \ell\rho$ , where  $\ell$  is constant, rewrite the traffic flow equation (5.34) in terms of  $\phi(x, t)$ .

## Section 5.5

**5.7** Solve the following problems by extending the method that was used in Sect. 5.5 to solve the advection equation.

(a)

$$\frac{\partial \rho}{\partial t} + 2 \frac{\partial \rho}{\partial x} = 1,$$

where  $\rho(x, 0) = 1/(1 + x^2)$  for  $-\infty < x < \infty$ .

(b)

$$\frac{\partial \rho}{\partial t} - 6 \frac{\partial \rho}{\partial x} = \rho,$$

where  $\rho(x, 0) = f(x)$  is given in (5.26).

(c)

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial x} = \rho^2,$$

where  $\rho(x, 0) = -1/(1 + x^2)$  for  $-\infty < x < \infty$ .

(d)

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial x} = x, \text{ for } x > 0, t > 0,$$

where  $\rho(x, 0) = 0$  for  $x > 0$ , and  $\rho(0, t) = 0$  for  $t > 0$ .

**5.8** The characteristics for the finite highway equation (5.31) are shown in Fig. 5.13. This problem explores other boundary and initial conditions that might be used.

- Suppose it is known that  $\rho(1, t) = 2$ . What is the initial condition  $\rho(x, 0)$  and what is the boundary condition  $\rho(0, t)$ ?
- Suppose it is known that  $\rho(1, t) = e^{-t}$ . What is the initial condition  $\rho(x, 0)$  and what is the boundary condition  $\rho(0, t)$ ?
- Suppose it is known that  $\rho(1, t) = 1/(1 + t)$ . What is the initial condition  $\rho(x, 0)$  and what is the boundary condition  $\rho(0, t)$ ?
- Suppose the solution is known at  $t = 5$ . For example, suppose  $\rho(x, 5) = x$ . Is it possible to determine the initial condition that was specified for the problem? Is it possible to determine the boundary condition at  $x = 0$  that was specified for the problem?

## Section 5.6

**5.9** In Fig. 5.19 it is assumed that  $0 < \rho_R < \rho_L < \frac{1}{2}\rho_M$ . In this exercise, some of the other possibilities are considered.

- Redraw Fig. 5.19 for the case of when  $0 < \rho_R < \frac{1}{2}\rho_M < \rho_L < \rho_M$ .
- Redraw Fig. 5.19 for the case of when  $\frac{1}{2}\rho_M < \rho_R < \rho_L < \rho_M$ .
- Redraw Fig. 5.19 for the case of when  $\rho_R = \frac{1}{2}\rho_M$  and  $\rho_L = \rho_M$ .

**5.10** This problem considers possible solutions of the traffic flow equation when using the Greenshields law, with  $v_M = 1$  and  $\rho_M = 6$ . Assume that  $\rho(x, 0) = f(x)$ , where  $f(x)$  is piecewise constant. Also, make sure to justify your answers.

- Give an example of  $f(x)$  that produces a solution with two expansion fans and no shock waves. Explain why the two expansion fans cannot overlap.
- Give an example of  $f(x)$  that produces a solution that starts out with two shock waves, both with negative slopes.
- Give an example of  $f(x)$  that produces a solution that starts out with a shock wave on the right, and an expansion fan on the left.
- In the example illustrated in Fig. 5.30, the shock and fan intersect. Is it possible to find an example where this does not happen?

**5.11** This problem considers what values a solution of the traffic flow equation can have when using the Greenshields law, with  $v_M = 1$  and  $\rho_M = 10$ . Assume that  $\rho(x, 0) = f(x)$  is piecewise constant.

- Assuming that  $3 \leq f(x) \leq 4$ , explain why it is impossible for  $\rho(x, t) = 5$  at any value of  $(x, t)$ . Is it possible for  $\rho(x, t) = 2$ ?
- Suppose that  $f(x)$  only takes on the values 1 and 3. Give an example where  $\rho(x, t)$  takes on all values between 1 and 3. Give an example where  $\rho(x, t)$  only takes on the values 1 and 3.
- Give an example for  $f(x)$ , so that the only values  $\rho$  takes on are 1, 2, and 3.

**5.12** This problem explores some of the connections between the velocity functions that arise with nonlinear traffic flow. Assume that  $v$  is a continuous function of  $\rho$ , for  $\rho \geq 0$ .

(a) Show that

$$v = \frac{1}{\rho} \int_0^\rho c(r) dr.$$

(b) Show that

$$s'(t) = \frac{\rho_R v_R - \rho_L v_L}{\rho_R - \rho_L}.$$

(c) If  $c_m \leq c(\rho) \leq c_M$ , for all possible  $\rho$ , show that  $c_m \leq s'(t) \leq c_M$ . What does this reduce to when the Greenshields law is used?

(d) If  $c(\rho) = v(\rho)$ , for all possible  $\rho$ , then what is  $v$ ?

(e) Is it possible for a shock wave to stay in one place? You can assume the Greenshields law is used.

(f) If the wave velocity  $c$  is independent of  $\rho$ , is the velocity  $v$  independent of  $\rho$ ?

**5.13** Find the solution of the traffic flow equation for the given initial condition, and then sketch the solution for  $t > 0$ . Use the Greenshields law with  $v_M = 2$  and  $\rho_M = 8$ .

- |                                                                                                                             |                                                                                                                                                                   |
|-----------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>(a) <math>\rho(x, 0) = \begin{cases} 1 &amp; \text{if } x &lt; 3 \\ 5 &amp; \text{if } 3 &lt; x \end{cases}</math></p>   | <p>(e) <math>\rho(x, 0) = \begin{cases} 1 &amp; \text{if } x &lt; -1 \\ 3 &amp; \text{if } -1 &lt; x &lt; 2 \\ 5 &amp; \text{if } 2 &lt; x \end{cases}</math></p> |
| <p>(b) <math>\rho(x, 0) = \begin{cases} 3 &amp; \text{if } x &lt; -1 \\ 1 &amp; \text{if } -1 &lt; x \end{cases}</math></p> | <p>(f) <math>\rho(x, 0) = \begin{cases} 0 &amp; \text{if } x &lt; 0 \\ 2 &amp; \text{if } 0 &lt; x &lt; 3 \\ 4 &amp; \text{if } 3 &lt; x \end{cases}</math></p>   |
| <p>(c) <math>\rho(x, 0) = \begin{cases} 4 &amp; \text{if } x &lt; 2 \\ 2 &amp; \text{if } 2 &lt; x \end{cases}</math></p>   | <p>(g) <math>\rho(x, 0) = \begin{cases} 4 &amp; \text{if } x &lt; 0 \\ 3 &amp; \text{if } 0 &lt; x &lt; 1 \\ 2 &amp; \text{if } 1 &lt; x \end{cases}</math></p>   |
| <p>(d) <math>\rho(x, 0) = \begin{cases} 5 &amp; \text{if } x &lt; 1 \\ 0 &amp; \text{if } 1 &lt; x \end{cases}</math></p>   |                                                                                                                                                                   |

**5.14** The two examples in Sect. 5.6.2 show that a continuous initial density  $\rho(x, 0) = f(x)$  can produce a solution that is shock free, or one that has a shock. In this problem assume that  $f(x)$  is smooth (so, its derivatives are defined and are continuous).

(a) What condition on  $f(x)$  guarantees that the resulting solution is shock free? Make sure to explain why.

- (b) Suppose that  $f(x) = 1/(1 + e^{\alpha x})$ . What restriction, if any, does your condition from part (a) put on  $\alpha$ ?

**5.15** This problem examines what happens on a finite length highway when the wave velocity is not constant. The equation is

$$\frac{\partial \rho}{\partial t} + c(\rho) \frac{\partial \rho}{\partial x} = 0, \quad \text{for } \begin{cases} 0 < x < \ell \\ 0 < t, \end{cases}$$

where

$$\rho(x, 0) = f(x), \quad \text{for } 0 \leq x \leq \ell,$$

and

$$\rho(0, t) = g(t), \quad \text{for } 0 < t.$$

Assume the Greenshields law is used, and so the function  $c$  is given in (5.37).

- Assuming that  $2\rho_0 < \rho_M$ , find the solution when  $f(x) = \rho_0(1 - x/\ell)$  and  $g(t) = \rho_0$ .
- For part (a), explain why there is no solution if  $2\rho_0 > \rho_M$ . Which condition, the one at  $t = 0$  or the one at  $x = 0$ , should be dropped so there is a solution for  $0 < x < \ell$  and  $t > 0$ ?
- Find the solution when  $f(x) = 0$  and  $g(t) = \rho_0$ , where  $0 < 2\rho_0 < \rho_M$ .
- Find the solution when  $f(x) = 0$  and  $g(t) = \rho_0$ , where  $\rho_M < 2\rho_0 \leq 2\rho_M$ .

**5.16** This problem explores what happens when a shock hits an expansion fan. The specific problem considered is shown in Fig. 5.30.

- Show that the shock hits the fan at  $t = 5$ . What is the value of  $s$  when this happens?
- Using the Rankine-Hugoniot condition show that as the shock passes through the fan,  $s' - \frac{1}{2t}s = \frac{3}{10} - \frac{1}{2t}$ .
- Solve the differential equation in part (b), and use part (a) to determine the integration constant.
- Show that the shock reaches the other side of the fan when  $t = 20$ .
- What is the equation of the shock once  $t > 20$ ?
- Similar to what is done in the lower figure in Fig. 5.30, sketch the solution when  $5 < t < 20$  and when  $20 < t$ .
- The upper figure in Fig. 5.30 is correct for  $0 \leq t \leq 5$ . Redraw the figure so it is correct for  $0 \leq t < \infty$ .

**5.17** To investigate just how much influence the constitutive law has on the solution, suppose it is assumed that  $v = v_M((1 - (\rho/\rho_M)^2))$ . This is a special case of what is known as Drew's constitutive law (Drew 1968).



- (a) What is the wave velocity  $c(\rho)$ ?
- (b) What is the solution of the modified red light–green light problem? As in Fig. 5.19, sketch the solution as a function of  $x$  and comment on how the solution differs from the one in Fig. 5.19. You can assume in this problem that  $\rho_R = 0$ .
- (c) What is the solution of the red light–green light problem, where  $\rho(x, 0)$  is given in (5.66)? Also, take  $v_M = 1$  and  $\rho_M = 4$ . As in Fig. 5.28, sketch the solution as a function of  $x$ . Also, comment on how the solution differs from the one in Fig. 5.28.
- (d) What is the solution of the traffic jam problem, where  $\rho(x, 0)$  is given in (5.57)? Also, take  $v_M = 1$  and  $\rho_M = 4$ . As in Fig. 5.23, sketch the solution as a function of  $x$ . Also, compare the velocity of the shock with the value obtained using the Greenshields law.

**5.18** It is observed that when a stoplight turns green, the flux of cars passing through the light increases in time up to a constant value  $J_0$ . Assuming the light is located at  $x = 0$ , a boundary condition that mimics this observed behavior is

$$J|_{x=0} = \begin{cases} J_0 t/T & \text{if } 0 \leq t \leq T \\ J_0 & \text{if } t > T, \end{cases}$$

The domain over which the traffic flow problem is solved is  $0 < x$  and  $0 < t$ . Assume here that  $\rho(x, 0) = 0$  and the Greenshields constitutive law is used. It is also assumed that  $0 < J_0 < J_c$ , where  $J_c = v_M \rho_M/4$  is the maximum flux possible and it is known as the capacity (see Exercise 5.5).

- (a) Rewrite the condition at  $x = 0$  in terms of  $\rho$ . Assume that  $0 \leq \rho \leq \rho_M/2$ .
- (b) What does the boundary condition in part (a) reduce to in the case of when  $J_0 = \frac{3}{4}J_c$ ?
- (c) Using the condition in part (b), solve the traffic flow problem.

**5.19** Suppose you had an experimental apparatus that enabled you to measure the velocity of a shock wave. Explain how you could use this to determine a constitutive law for the velocity.

**5.20** Suppose the density is given in terms of the velocity, and so, assume  $\rho = H(v)$ .

- (a) Show how the traffic flow equation can be written as  $v_t + d(v)v_x = 0$ .
- (b) Find  $H$  for the Greenshields (5.10) and Newell (5.17) functions.
- (c) The initial condition used for the small disturbance approximation is  $v(x, 0) = v_0 + \varepsilon h(x)$ . Find the resulting two term expansion for the velocity.

**5.21** One might argue that if a driver is in a relatively high-density region and sees lower density traffic up ahead that they will speed up with the objective of traveling in the lower-density region.

- (a) Explain why an assumption that accounts for this behavior is a constitutive law of the form  $v = F(\rho, \rho_x)$ .
- (b) Write down a simple, three parameter, constitutive law for  $v$  that involves  $\rho$  and  $\rho_x$ . Your law should reduce to the Greenshields law when  $\rho_x = 0$ .
- (c) With the constitutive law from part (b), what is the resulting traffic flow equation?
- (d) What is the resulting small disturbance equation and how does it differ from (5.41)?

## Section 5.7

**5.22** This problem considers various formulas for the NaSch model.

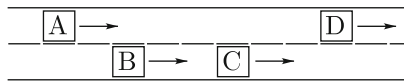
- (a) Show that the first two steps in the NaSch model can be combined into the formula  $m_{new} = \min\{m_{old} + 1, M, g_{old}\}$ .
- (b) Show that the first three steps in the NaSch model can be combined into the formula  $m_{new} = \max\{0, \min\{m_{old} + 1, M, g_{old}\} - \chi\}$ , where  $\chi = 1$  with probability  $p$  (otherwise,  $\chi = 0$ ).

**5.23** Suppose that in the NaSch model the cars start out uniformly spaced with  $g = M$ . Assume the randomization probability is  $p = 1$ .

- (a) Let  $M = 1$ . What happens if the cars start out with velocity  $M$ ? What happens if the cars start out with zero velocity?
- (b) Let  $M = 3$ . What happens if the cars start out with velocity  $M$ ? What happens if the cars start out with velocity  $m = 2$ ? What happens if the cars start out with velocity  $m = 1$ ? What happens if the cars start out with zero velocity?
- (c) Generalize your conclusions from part (b) to describe what happens if  $M \geq 2$ .

**5.24** Suppose there is a stoplight located at  $x = 0$ . When it turns red assume the cars are uniformly spaced in the region  $x < 0$ , with three spaces between the cars, and each car has  $m = 1$ . The maximum movement is  $M = 2$ . A space here is one car length.

- (a) Assuming the Greenshields constitutive law is used, what is the resulting solution of the traffic flow problem? What is the velocity of the shock wave?
- (b) In the NaSch model suppose the randomization is turned off (i.e.,  $p = 0$ ). Show that the approximate velocity of the shock wave is  $-2\Delta x/(3\Delta t)$ . How does this compare with the continuum result from part (a)?
- (c) In the NaSch model suppose that in the randomization step  $p = 1$ . Explain why there is a shock-like solution but the jam density is half of what is obtained from the continuum solution. Also show that the shock moves with approximate velocity  $-\Delta x/\Delta t$ .
- (d) Given the solution in part (c) describe, in general terms, what happens when  $p$  is close to one.



**Fig. 5.39** Possible car positions when deciding to make a lane change, as considered in Problem 5.25

- (e) Using your results from parts (b) and (d), explain why  $-2(1 - p)\Delta x/(3\Delta t)$  provides an approximation of the shock velocity for the NaSch model. Given this, what should the randomization probability be so that the NaSch velocity agrees with the continuum model?

**5.25** To extend the NaSch model to multilane roads, where individual cars are able to change lanes, consider Fig. 5.39. Assume a driver will switch lanes whenever they are able to travel farther in a time step in the other lane. Safe lane changing requires consideration of the backward gap in the other lane, so the driver in car B must consider the position and velocity of car A when deciding to switch. Write down a set of rules for moving the cars along the highway that includes lane changes. Assume in this problem that the randomization step is omitted.

### Additional Questions

**5.26** In fluid dynamics one across the problem of solving

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0,$$

which is known as Burgers' equation and  $u(x, t)$  is the velocity of the fluid. In this case,  $u$  can be positive or negative.

- (a) What is the wave velocity version of the Rankine-Hugoniot condition for this problem? What is the flux  $J$ ?  
 (b) Suppose the initial condition is

$$u(x, 0) = \begin{cases} u_L & \text{if } x < 2 \\ u_R & \text{if } 2 < x. \end{cases}$$

When will a shock be generated, and when will there be an expansion fan? Explain how this differs for the traffic flow problem, when the Greenshields law is used.

(c) Assume that the initial condition is

$$u(x, 0) = \begin{cases} 2 & \text{if } x < -1 \\ -1 & \text{if } -1 < x. \end{cases}$$

Sketch the characteristics, and use this to find the solution.

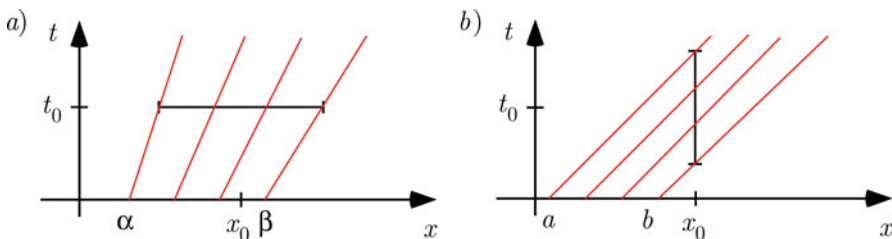
(d) Assume that the initial condition is

$$u(x, 0) = \begin{cases} -1 & \text{if } x < -1 \\ 2 & \text{if } -1 < x. \end{cases}$$

Sketch the characteristics, and use this to find the solution.

**5.27** This problem examines the averaging used to define the flux and density, and how they relate with the velocity. It is assumed that a car with initial location  $x_0$  has velocity  $f(x_0)$ . Consequently, the position of this car at any later time  $t$  is  $x = x_0 + f(x_0)t$ . Example paths for the cars are shown in Fig. 5.40. Therefore, in this problem, each car has a constant velocity, but different cars can have different velocities. For simplicity, it is assumed that  $f(x) = v_0 + w_0x$ , where  $v_0$  and  $w_0$  are constants.

- The averaging interval used to define the density in (5.1) is shown in Fig. 5.40a, and it is the same as the one shown in Fig. 5.4. Explain why  $x_0 - \Delta x = \alpha + f(\alpha)t_0$  and  $x_0 + \Delta x = \beta + f(\beta)t_0$ . Use these equations to find  $\alpha$  and  $\beta$  in terms of  $x_0$  and  $t_0$ .
- The averaging interval used to define the flux in (5.3) is shown in Fig. 5.40b. Explain why  $t_0 - \Delta t = b + f(b)t_0$  and  $t_0 + \Delta t = a + f(a)t_0$ . Use these equations to find  $a$  and  $b$  in terms of  $x_0$  and  $t_0$ .
- Assuming that the cars are continuously distributed, show that the average velocity for the cars in the horizontal bar in Fig. 5.40a is  $v_0 + w_0(x_0 - v_0t_0)/(1 + w_0t_0)$ .



**Fig. 5.40** Averaging intervals used to define (a) the density, and (b) the flux. The horizontal bar in (a) has length  $2\Delta x$ , and the vertical bar in (b) has length  $2\Delta t$ . The four slanted lines shown in each figure are the paths of individual cars. These figures are used in Problem 5.27

- (d) Assuming that the cars are continuously distributed, find the average velocity for the cars in the vertical bar in Fig. 5.40b. Assuming that  $\Delta t$  is small, show that the average velocity is, approximately,  $v_0 + w_0(x_0 - v_0 t_0)/(1 + w_0 t_0)$ .
- (e) Use the fact that the average velocities in parts (c) and (d) are the same to explain why this provides additional evidence of the validity of the equation  $J = \rho v$ .

**5.28** One way to explain weak solutions is to consider a smooth version of the jump initial condition. Specifically, let  $\rho(x, 0) = 1/(1 + \alpha e^{x/\varepsilon})$ , where  $\varepsilon$  and  $\alpha$  are positive. Also, this problem considers the linear equation, so  $v = a$  is constant.

- (a) Find the solution using the above initial condition. Explain why the resulting solution is smooth and sketch it assuming that  $\varepsilon$  is small.
- (b) Explain what happens both to the initial condition and solution when  $\varepsilon \rightarrow 0$ . Make sure to explain what happens if  $x = at$ .
- (c) Part (b) helps explain why using a jump initial condition is consistent with what is known for smooth solutions, with the exception of what happens at the jump itself. This raises the question of what value the density can have at a jump. Given the definition of the density in (5.1), what should the value of the density be at  $x = 0$  and at  $x = 1$  in Fig. 5.9 when  $t = 0$ ?
- (d) Show that the limiting value you found in part (b), when  $x = at$ , can be obtained for the solution in Fig. 5.9 by modifying the averaging interval in (5.1). What this shows is that in a continuum theory, the value at a discontinuity is dependent on the averaging method.

**5.29** This problem examines what happens to the traffic flow problem when cars are allowed to enter or exit the highway. It is assumed this occurs not at discrete locations but continuously along the highway.

- (a) Assume that over an interval  $x_0 - \Delta x < x < x_0 + \Delta x$  the number that enter (or exit) from  $t = t_0 - \Delta t$  to  $t = t_0 + \Delta t$  is  $4\Delta x \Delta t Q$ , where  $Q(x, t)$  is the net rate per unit length at which cars are entering or leaving the highway. Show that the resulting balance law for traffic flow is

$$\frac{\partial \rho}{\partial t} = -\frac{\partial J}{\partial x} + Q.$$

- (b) One possible constitutive law for this new variable is  $Q = \alpha(\rho - \beta)$  where  $\alpha, \beta$  are constants. Can you explain how this assumption could arise for traffic flow? Is there any reason you should assume  $\alpha$  is either positive or negative? Any suggestion on how to choose  $\beta$ ?
- (c) Use the procedure to solve the  $\alpha = 0$  case to solve the equation derived in part (a) along with the constitutive assumption in part (b). Assume a constant velocity.
- (d) Based on your solution from part (c), what is the effect of  $Q$  on the density? Is the solution still a traveling wave? Demonstrate your conclusion using the initial distribution  $\rho(x, 0) = e^{-x^2}$  by sketching the solution at later times.

**5.30** This problem investigates how to use similarity variables to find an expansion fan.

- Assume  $\rho(x, t) = R(\eta)$ , where  $\eta = x/t$ . Show that the traffic flow equation reduces, in the case of when  $\rho$  is not constant, to the equation  $c(\rho) = x/t$ .
- Using the Greenshields law, solve  $c(\rho) = x/t$  for  $\rho$ .
- Show that your solution in part (b) is the same as the one given in (5.62).

**5.31** Suppose one is interested in knowing the position of a particular car when using the continuum model. Assume that at  $t = 0$  the car is located at  $x = A$ , and its position at later times is given by  $x = \chi(t)$ . This problem is concerned with how to find the function  $\chi(t)$ . In doing this it is assumed that the traffic flow equation has been solved, so the density  $\rho(x, t)$  and the velocity  $v(x, t)$  functions are known.

- Explain why, to find  $\chi(t)$ , one solves the differential equation  $\chi' = v(\chi, t)$ , where  $\chi(0) = A$ .
- For the red light-green light solution given in (5.67), what is  $v(x, t)$ ? For this velocity function, and assuming that  $A < 0$ , solve the problem in part (a) and show that

$$\chi(t) = \begin{cases} A & \text{if } 0 \leq t \leq -A/2 \\ 2(t - \sqrt{-2At}) & \text{if } -A/2 < t. \end{cases}$$

On the same axes, sketch  $\chi(t)$  for  $A = -2$ ,  $A = -4$ , and  $A = -6$ .

- For the red light-green light problem, which cars are able to get through the light if it is green for  $0 \leq t < t_R$  and turns red at  $t = t_R$ ?
- For the traffic jam example studied in Sect. 5.6.4, find  $\chi(t)$  for  $A < 0$ . On the same axes, sketch  $\chi(t)$  for  $A = -1$ ,  $A = -2$ , and  $A = -3$ .

**5.32** This problem considers how a traffic flow equation can be derived from a particle model similar to how the diffusion equation can be derived from a random walk (see Sect. 4.2). It is assumed that the particles can occupy positions  $x = m\Delta x$  on the  $x$ -axis, where  $m = 0, \pm 1, \pm 2, \dots$ . In going from  $t = (n-1)\Delta t$  to  $t = n\Delta t$ , a particle located at  $m\Delta x$  will jump to  $(m+1)\Delta x$  if  $(m+1)\Delta x$  is unoccupied at  $t = (n-1)\Delta t$ , otherwise the particle will not move. Letting  $w(m, n)$  be the occupation function, then  $w(m, n) = 1$  if a particle is located at  $x = m\Delta x, t = n\Delta t$ , otherwise  $w(m, n) = 0$ . The model being considered here is a version of what is called an asymmetric simple exclusion process (ASEP) and it arises in statistical mechanics (Blythe and Evans, 2007; Golinelli and Mallick, 2006) and biophysics (Reese et al., 2011).

- Suppose at  $n = 0$  there are five particles and they are at  $m = 0, m = 3, m = 4, m = 5$ , and  $m = 10$ . Where are they located at  $n = 3$ ?
- There are two configurations at  $n = 1$  that can result in  $w(m, n) = 1$ . What are they? There are also two configurations at  $n = 1$  that can result in  $w(m, n) = 0$ . What are they?

(c) The resulting master equation is

$$w(m, n) = [1 - w(m, n - 1)]w(m - 1, n - 1) + w(m, n - 1)w(m + 1, n - 1).$$

Show that this gives the correct value for the four configurations in part (b).

- (d) Letting  $u(x, t)$  be the continuum version of the occupation function, derive the resulting partial differential equation for  $u(x, t)$ . Make sure to state the requirement imposed on  $\Delta x$  and  $\Delta t$ . Because of the application being considered, it is assumed that  $0 \leq u(x, t) \leq 1$ .
- (e) For the equation derived in part (d), explain why the corresponding flux is  $J(u) = v_0 u(1 - u)$  where  $v_0$  is a positive constant. Explain how this is connected with the Greenshields law.

**5.33** This problem explores the nonuniqueness of the traffic flow problem, and the importance of knowing the physical problem that it is associated with. Assume the Greenshields law holds, with  $v_M = 1$  and  $\rho_M = 2$ .

- (a) What does the Rankine-Hugoniot condition (5.52) give for the velocity of a shock?
- (b) Multiple the traffic flow equation (5.34) by  $\rho$  and show that the resulting equation can be written as  $\partial_t(\frac{1}{2}\rho^2) + \partial_x J = 0$ , where  $J = \frac{1}{2}\rho^2 - \rho^3$ .
- (c) For the equation you derived in part (b), use the same argument used in Sect. 5.6.3 to show that

$$s'(t) = \frac{2}{\rho_R + \rho_L} \frac{J(\rho_R) - J(\rho_L)}{\rho_R - \rho_L}.$$

Explain why this is generally different than the velocity you obtained in part (a).

- (d) Redo parts (b) and (c), but multiply by  $\rho^n$ , where  $n$  is a positive integer.

Comment: This problem demonstrates that there are an infinite number of shock velocities that are mathematically consistent with the original traffic flow equation. There is, however, only one that is derived from the conservation law that was used to obtain the traffic flow equation in Sect. 5.3.

# Chapter 6

## Continuum Mechanics: One Spatial Dimension



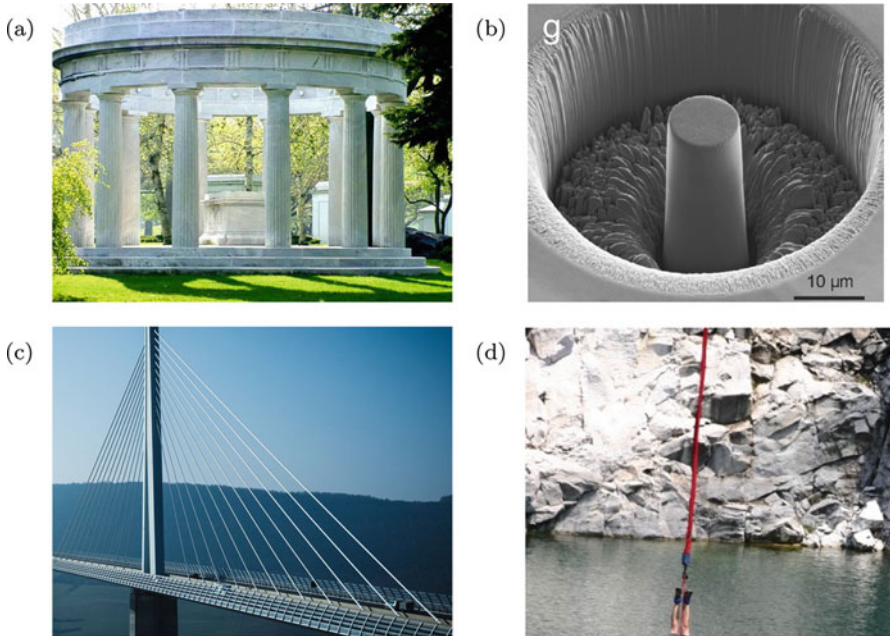
### 6.1 Introduction

In the previous chapter we investigated how to model the spatial motion of objects (cars, molecules, etc.) but omitted the possibility that the objects exert forces on each other. The objective now is to introduce this into the modeling. The situations where this is needed are quite varied and include the deformation of an elastic bar, the stretching of a string, or the flow of air or water. Each of these has an internal material structure that resists either stretching (the string and bar) or compression (air, water, and bar). Illustrations of particular applications are shown in Fig. 6.1. In this chapter a mathematical model for these systems is derived, and in doing this we will limit our attention to one-dimensional motion. Also, the only problems solved in this chapter are to find the steady-state solution. The question of how to solve the time-dependent problem coming from the model will be taken up in the next chapter.

### 6.2 Frame of Reference

The first step in continuum modeling is to specify the frame of reference that will be used. Before defining this mathematically, it is worth revisiting the traffic flow problem considered in the last chapter. From a driver's point of view, what we will call the material frame of reference, the driver knows how far they have traveled, their current position, their velocity, and the density of traffic in their vicinity. In comparison, someone standing on the side of the road, what we will call the spatial frame of reference, will know the driver's current position, their velocity, and the density of traffic. Both observers will agree on the values for the position, velocity and density. Where they differ is that the person on the side of the road will not know





**Fig. 6.1** Examples of uniaxial motion or deformation. (a) Doric-style columns that are under compression due to the annular entablature they support. (b) A micropillar of austenitic stainless steel used in materials with a composite microstructure (Guo et al., 2014). (c) Tension cables used to support the Millau Viaduct. (d) Longitudinal stretching of bungee cord

where the driver came from or how far they have traveled. Both frames of reference are very useful, and we will continually switch back and forth between them.

### 6.2.1 Material Coordinates

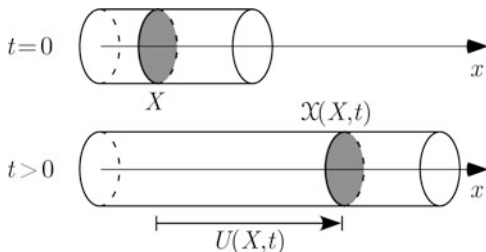
To define this system, consider a cylindrical bar as shown in Fig. 6.2. The top figure shows the bar at  $t = 0$  and identifies a particular cross-section located at  $x = X$ . The lower figure shows the bar at a later time and the cross-section has moved to  $x = \mathcal{X}(X, t)$ . To be consistent it is required that  $\mathcal{X}(X, 0) = X$ .

Given the position function  $\mathcal{X}$ , we can track each cross-section as it moves back and forth along the  $x$ -axis. In this way, the coordinates  $X, t$  can be used to locate the cross-sections, and together they constitute what is known as the material coordinate system. In physics this is usually referred to as Lagrangian coordinates.

We are particularly interested in how far each section moves from its original position, and for this reason we introduce the *displacement* function, defined as

$$U(X, t) \equiv \mathcal{X}(X, t) - X. \quad (6.1)$$

**Fig. 6.2** Axial motion of a cross-section that begins at  $x = X$  and moves to  $x = \mathcal{X}(X, t)$ . The resulting displacement of the cross-section is  $U(X, t)$



Because  $\mathcal{X}(X, 0) = X$ , then it is required that

$$U(X, 0) = 0. \quad (6.2)$$

Connected with the displacement is the *velocity* of the cross-section, defined as

$$V(X, t) \equiv \frac{\partial U}{\partial t}. \quad (6.3)$$

The acceleration and other higher time derivatives can be calculated similarly.

One last comment to make is that in the material coordinate system the positions of the cross-sections at  $t = 0$  comprise what is known as the reference configuration.

### 6.2.2 Spatial Coordinates

For an external observer watching the motion, the material description is not particularly convenient. For example, if you were to observe fluid flow in a river it is likely you would simply stand on the river bank and take your measurements at that fixed position. In doing this, you would be determining the properties of the motion without knowing where the individual water molecules were located at  $t = 0$ . For this reason the spatial, or Eulerian, coordinate system is introduced. The idea here is that you select the spatial location  $x$  and let the cross-sections come to you. In spatial coordinates the displacement function is denoted as  $u(x, t)$  and the velocity is  $v(x, t)$ . In this context,  $u(x, t)$  is the displacement of the cross-section that is located at  $x$  and time  $t$ , and  $v(x, t)$  is the velocity of that cross-section. So, at a fixed location  $x$ , as long as the velocity is nonzero, the cross-sections at  $x$  keep changing. In contrast, in the material system, a fixed  $X$  means you are following a particular cross-section, but your spatial location is changing.

The values obtained from the spatial coordinate system must be exactly the same as obtained when they are computed using material coordinates. To determine the consequences of this, the transformation between the material and spatial coordinates is through the formula  $x = \mathcal{X}(X, t)$ . Because the two coordinate systems must give the same value, we have that

$$U(X, t) = u(\mathcal{X}(X, t), t), \quad (6.4)$$

and

$$V(X, t) = v(\mathcal{X}(X, t), t). \quad (6.5)$$

If, on the other hand, we know the material variables  $U$  and  $V$  and want to calculate the spatial versions, it is first necessary to solve  $x = \mathcal{X}(X, t)$  for  $X$ . Physically this corresponds to determining the original position  $X$  of the cross-section that is currently located at  $x$ . Writing this as  $X = \mathcal{X}(x, t)$ , then

$$u(x, t) = U(\mathcal{X}(x, t), t), \quad (6.6)$$

and

$$v(x, t) = V(\mathcal{X}(x, t), t). \quad (6.7)$$

The above expressions will prove to be invaluable when transforming between spatial to material coordinates.

*Example* Suppose a cross-section that started at  $x = 3$  is, at  $t = 2$ , located at  $x = 7$ .

*Material Coordinates:* For this cross-section,  $X = 3$ , and its displacement at  $t = 2$  is  $7 - 3 = 4$ . In other words,  $U(3, 2) = 4$ . We also have that  $\mathcal{X}(3, 0) = 3$ , and  $\mathcal{X}(3, 2) = U(3, 2) + X = 7$ .

*Spatial Coordinates:* At  $t = 2$ , the displacement of the cross-section located at  $x = 7$  is 4. In other words,  $u(7, 2) = 4$ . ■

*Example* Suppose the bar deforms in such a way that the cross-sections move according to the rule that  $\mathcal{X} = X(1 + 2t)/(1 + t)$ . In this case,

$$U(X, t) = \mathcal{X} - X = \frac{tX}{1 + t},$$

and

$$V(X, t) = \frac{\partial U}{\partial t} = \frac{X}{(1 + t)^2}.$$

This means, for example, that the cross-section that starts at  $X = 1$  moves with velocity  $V = 1/(1 + t)^2$ . To express these formulas in spatial coordinates, we solve the rule to obtain  $X = x(1 + t)/(1 + 2t)$ . In this case, from (6.6),

$$u(x, t) = U \Big|_{X=\frac{1+t}{1+2t}x} = \frac{tx}{1 + 2t},$$

and, from (6.7),

$$v(x, t) = V \Big|_{X=\frac{1+t}{1+2t}x} = \frac{x}{(1+2t)(1+t)}. \quad (6.8)$$

One significant observation from the above calculation is that  $v \neq \frac{\partial u}{\partial t}$ . Also, the change of variables is reversible, and so

$$\begin{aligned} U(X, t) &= u(x, t) \\ &= u\left(X \frac{1+2t}{1+t}, t\right) \\ &= \frac{tX}{1+t}. \end{aligned} \quad \blacksquare$$

One of the distinctive differences in the two coordinate systems is the domain over which they apply. To demonstrate, suppose that the bar starts off, at  $t = 0$ , occupying the interval  $0 \leq X \leq \ell_0$ . The range of  $X$  values for any material variable, such as  $U(X, t)$  or  $V(X, t)$ , is determined by the original position of the bar. In other words, they are defined for  $0 \leq X \leq \ell_0$ . In contrast, the range of  $x$  values for a spatial variable, such as  $u(x, t)$  or  $v(x, t)$ , depends on the displacement of the bar. The left end of the bar is at  $x = \mathcal{X}(0, t) = U(0, t)$  and the right end is at  $x = \mathcal{X}(\ell_0, t) = \ell_0 + U(\ell_0, t)$ . Consequently, the spatial variables are defined for  $U(0, t) \leq x \leq \ell_0 + U(\ell_0, t)$ .

*Example* In the previous example, suppose the bar initially occupies the interval  $0 \leq X \leq 3$ . In this case,  $U(X, t)$  and  $V(X, t)$  are defined, for all values of  $t$ , in the interval  $0 \leq X \leq 3$ . To determine the interval for the spatial coordinates, note that the position of the left end is at

$$x = U(0, t) = 0,$$

and the position of the right end is at

$$x = 3 + U(3, t) = 3 + \frac{3t}{1+t}.$$

Therefore, the spatial variables are defined for  $0 \leq x \leq 3(1+2t)/(1+t)$ . ■

A comment is needed about partial derivatives and the two coordinate systems. The independent variables in the material system are  $X$  and  $t$ , while the independent variables in the spatial system are  $x$  and  $t$ . Partial derivatives with respect to these variables will be written using one of several notations. In particular, the following are three different ways to designate the first partial derivative:

$$\frac{\partial U}{\partial X}, \quad U_X, \quad \partial_X U.$$

The second partial derivatives are written in either of the following three ways:

$$\frac{\partial^2 U}{\partial X^2}, \quad U_{XX}, \quad \partial_X^2 U.$$

Correspondingly, mixed second partial derivatives are written in one of the following ways:

$$\frac{\partial^2 U}{\partial X \partial t}, \quad U_{Xt}, \quad \partial_X \partial_t U.$$

There is nothing particularly special about any one of these forms, and the choice of what to use is mostly determined by the format of the mathematical expression under consideration.

### 6.2.3 Material Derivative

The above example shows that even though  $V = \frac{\partial u}{\partial t}$  it turns out that, in general, in spatial coordinates  $v \neq \frac{\partial u}{\partial t}$ . This begs the question as how it might be possible to determine  $v$  if we know  $u$ . To answer this suppose  $F(X, t)$  is a function in material coordinates, and assume that its spatial version is  $f(x, t)$ . In this case  $\frac{\partial F}{\partial t}$  represents the time rate of change of the variable for the cross-section that began at  $X$ . To determine what this is in spatial coordinates note that  $F$  and  $f$  must produce the same value. For example, if the cross-section that started at  $X$  is currently located at  $x = \mathcal{X}(X, t)$ , then  $F(X, t) = f(x, t)$ . In other words,

$$F(X, t) = f(\mathcal{X}(X, t), t). \quad (6.9)$$

Taking the time derivative, and using the chain rule, we have that

$$\begin{aligned} \frac{\partial F}{\partial t} &= \frac{\partial f}{\partial x} \frac{\partial \mathcal{X}}{\partial t} + \frac{\partial f}{\partial t} \\ &= \frac{\partial f}{\partial x} V(X, t) + \frac{\partial f}{\partial t} \\ &= \frac{\partial f}{\partial x} v(x, t) + \frac{\partial f}{\partial t}. \end{aligned} \quad (6.10)$$

This derivative plays such an important role in what follows that it gets its own name and notation.

**Material Derivative.** *The material derivative, which is defined as*

$$\frac{Df}{Dt} \equiv \frac{\partial f}{\partial t} + v \frac{\partial f}{\partial x}, \quad (6.11)$$

*is the time rate of change of a function following a material cross-section, but expressed in spatial coordinates.*

A particularly important application of the above formula concerns the spatial coordinate description of the displacement and velocity functions. Taking  $F(X, t) = U(X, t)$ , then  $f(x, t) = u(x, t)$  and from (6.10) we have that

$$\frac{\partial U}{\partial t} = \frac{Du}{Dt}. \quad (6.12)$$

Because  $V = \frac{\partial U}{\partial t}$ , and  $V(X, t) = v(x, t)$ , it follows that

$$v = \frac{Du}{Dt}. \quad (6.13)$$

Using the definition of the material derivative in (6.11) the above equation reduces to

$$v = \frac{u_t}{1 - u_x}. \quad (6.14)$$

In other words, the velocity  $v(x, t)$  of the cross-section located at  $x$ , at time  $t$ , is  $\frac{Du}{Dt}$  and not  $\frac{\partial u}{\partial t}$ . This result explains why in spatial coordinates we usually end up with  $v \neq u_t$ .

*Example* Using the functions from the previous example,

$$\begin{aligned} \frac{\partial u}{\partial x} &= \frac{t}{1 + 2t}, \\ \frac{\partial u}{\partial t} &= \frac{x}{(1 + 2t)^2}, \end{aligned}$$

and so, from (6.14)

$$v = \frac{x}{(1 + t)(1 + 2t)}.$$

This agrees with the formula for the spatial velocity calculated by converting the material velocity  $V$  to  $v$  in (6.8). ■

In modeling the deformation of a bar we will need to consider the relative changes in the material. The derivative of interest in such situations is the material

**Table 6.1** Relationship between various spatial and material derivatives

Material Expression		Spatial Expression
$\frac{\partial F}{\partial t}$	$=$	$\frac{\partial f}{\partial t} + v \frac{\partial f}{\partial x}$
$\frac{\partial F}{\partial X}$	$=$	$\frac{1}{1 - u_x} \frac{\partial f}{\partial x}$
$\frac{1}{1 + U_X} \frac{\partial F}{\partial X}$	$=$	$\frac{\partial f}{\partial x}$
$\frac{\partial F}{\partial t} - \frac{V}{1 + U_X} \frac{\partial F}{\partial X}$	$=$	$\frac{\partial f}{\partial t}$

Here  $F(X, t)$  is a function in material coordinates and  $f(x, t)$  is the function in spatial coordinates.

gradient  $\frac{\partial F}{\partial X}$ . To express this in spatial coordinates we again start by differentiating (6.9), but this time with respect to  $X$ . In a similar manner as before, using the chain rule one finds that

$$\begin{aligned} \frac{\partial F}{\partial X} &= \frac{\partial f}{\partial x} \frac{\partial x}{\partial X} \\ &= \frac{\partial f}{\partial x} \left( 1 + \frac{\partial U}{\partial X} \right). \end{aligned} \quad (6.15)$$

In the special case of when  $F = U$  and  $f = u$ , this reduces to  $U_X = (1 + U_X)u_x$ . Solving for  $U_X$  we obtain

$$\frac{\partial U}{\partial X} = \frac{u_x}{1 - u_x}. \quad (6.16)$$

Substituting this into (6.15), yields

$$\frac{\partial F}{\partial X} = \frac{1}{1 - u_x} \frac{\partial f}{\partial x}. \quad (6.17)$$

This result is listed in Table 6.1 along with other useful equalities between various derivatives.

### 6.2.4 End Notes

Before moving on to the derivation of the equations of motion there are a few aspects of the model that need to be stated explicitly. First, it is important to remember that the points making up a material cross-section, such as shown in Fig. 6.2, move as a unit and the motion is only along the  $x$ -axis. One consequence of this assumption is that the cross-sectional area is constant.

A second point is the implicit assumption that cross-sections do not pass each other. For example, if a cross-section starts out to the left of another cross-section, then it is always to the left of it. In mathematical terms, this assumption can be written as

$$\mathcal{X}(X_1, t) < \mathcal{X}(X_2, t), \quad \forall X_1 < X_2. \quad (6.18)$$

This is reasonable from a physical viewpoint, and it is known as the *impenetrability of matter assumption*. As will be explained in the next section, this inequality guarantees that we can uniquely convert back and forth between spatial and material coordinates. To guarantee that (6.18) holds it is assumed that (see Exercise 6.4)

$$\frac{\partial U}{\partial X} > -1. \quad (6.19)$$

This condition is equivalent to the assumption that (see Exercise 6.4)

$$\frac{\partial u}{\partial x} < 1. \quad (6.20)$$

A third implicit assumption being made concerns the smoothness of the motion. We will differentiate variables any number of times based on the assumption that the motion is smooth enough to permit this.

There is more than a passing connection between this chapter and the previous one on traffic flow. In a mathematical sense the cars of the last chapter are cross-sections in this chapter. Both are objects moving along the  $x$ -axis. For this reason it should not be a surprise that the cross-sections satisfy a conservation law related to their density, just as the cars did in the last chapter as expressed in (5.7). A significant difference is that in this chapter the objects (i.e., the cross-sections) exert forces on their neighbors, and we will derive a force balance equation from this. Another difference is that only a spatial coordinate system was used in traffic flow. It was not necessary to introduce the material coordinate system, which would describe the motion from the driver's point of view, but in this chapter the material system is important. To explain why, using the car analogy, we will assume that the car directly in front of the driver, and the car directly behind, exert a force on the driver's car. This will happen, for example, if adjacent cars were connected by springs. These forces change as the distance between the cars change, and to keep track of this we will need to follow the cars. Hence, the need for material coordinates. At the same time, there will be situations where the spatial coordinate system is preferable, and this means we will switch back and forth between the two systems. If you are interested how material coordinates can be used in traffic flow, Exercise 5.31 should be consulted.



### 6.3 Mathematical Tools

To derive the traffic flow equation in the last chapter we used a control volume approach. With the objective of trying different ideas, we will derive the equations in this chapter using what is known as the integral method. This requires a bit more mathematical background but the benefit is that the derivation is easier. It is the purpose of this section to present the needed mathematical tools. As with seemingly everything in mathematical modeling, these results are personalized and given names.

The first result is straight out of mathematical analysis, and it tells us when we are able to conclude a function is identically zero.

**du Bois-Reymond Lemma.** *If  $f(x)$  is continuous and  $\int_a^b f(x)dx = 0$ ,  $\forall a, b$  with  $a < b$ , then  $f(x) = 0$ ,  $\forall x$ .*

The usual proof of this result involves contradiction. Assuming there is a point where  $f(x) \neq 0$ , then continuity requires that there is a small interval where the function is either positive or negative. The existence of such an interval contradicts the zero integral assumption and therefore such a point cannot exist. Just as a note in passing, the fellow this result is named for, Paul du Bois-Reymond, was the brother of Emil, the noted physiologist.

The second result we need involves the rate of change of an integral in which the interval of integration depends on time.

**Reynolds Transport Theorem.** *Assuming  $\alpha(t)$ ,  $\beta(t)$ ,  $f(x, t)$  are smooth functions, then*

$$\frac{d}{dt} \int_{\alpha(t)}^{\beta(t)} f(x, t) dx = \int_{\alpha(t)}^{\beta(t)} \frac{\partial f}{\partial t} dx + f(\beta, t)\beta'(t) - f(\alpha, t)\alpha'(t). \quad (6.21)$$

In looking at this result it might appear we have simply restated a version of Leibniz's rule for differentiation under the integral sign. This observation is correct. The Reynolds Transport Theorem, as usually stated, is for the time rate of change of a volume integral when the integration domain is time dependent. Our version is what is obtained when reducing the motion to the  $x$ -axis.

Given that we are establishing the mathematical tools for the derivation of the equations of motion it is appropriate to mention the inverse function theorem. In transforming between the material and spatial systems we need to solve  $x = \mathcal{X}(X, t)$  for  $X$ . The conditions needed to guarantee that this is possible are contained in the next theorem.

**Inverse Function Theorem.** *Given a closed interval  $I$ , and a continuous function  $F$  that is strictly monotonic on  $I$ , then  $F(I)$  is a closed interval and the inverse function  $F^{-1}$  is defined, strictly monotonic, and continuous on  $F(I)$ .*

The impenetrability of matter assumption (6.18) is an assumption of strict monotonicity, and so we can uniquely transform back and forth between the material and spatial systems.

## 6.4 Continuity Equation

We will assume that mass is neither created nor destroyed. To understand the mathematical consequences of this assumption suppose at  $t = 0$  we identify a segment of the bar, and assume this is an interval of the form  $X_L \leq x \leq X_R$ . At any later time this segment occupies an interval  $\alpha(t) \leq x \leq \beta(t)$ , where the endpoints are determined from the formulas  $\alpha(t) = \mathcal{X}(X_L, t)$  and  $\beta(t) = \mathcal{X}(X_R, t)$ . Our assumption means that the total mass of the material in this interval does not change. If we let  $\rho(x, t)$  designate the mass density of the material (i.e., mass per unit volume), then our assumption states that

$$\int_{\alpha(t)}^{\beta(t)} \sigma \rho(x, t) dx = \int_{X_L}^{X_R} \sigma \rho(x, 0) dx, \quad (6.22)$$

where  $\sigma$  is the (constant) cross-sectional area of the bar. Differentiating both sides with respect to  $t$  gives

$$\frac{d}{dt} \int_{\alpha(t)}^{\beta(t)} \sigma \rho(x, t) dx = 0. \quad (6.23)$$

Recalling that  $\partial_t \mathcal{X}(X, t) = V(X, t)$  and  $V(X, t) = v(\mathcal{X}(X, t), t)$ , then  $\beta'(t) = v(\beta, t)$  and  $\alpha'(t) = v(\alpha, t)$ . With this, and the Reynolds Transport Theorem, we have the following

$$\begin{aligned} & \frac{d}{dt} \int_{\alpha(t)}^{\beta(t)} \sigma \rho(x, t) dx \\ &= \int_{\alpha(t)}^{\beta(t)} \sigma \frac{\partial \rho}{\partial t} dx + \sigma \rho(\beta, t) \beta' - \sigma \rho(\alpha, t) \alpha' \end{aligned} \quad (6.24)$$

$$\begin{aligned} &= \int_{\alpha(t)}^{\beta(t)} \sigma \frac{\partial \rho}{\partial t} dx + \sigma \rho v|_{x=\beta} - \sigma \rho v|_{x=\alpha} \\ &= \int_{\alpha(t)}^{\beta(t)} \sigma \frac{\partial \rho}{\partial t} dx + \int_{\alpha(t)}^{\beta(t)} \sigma \frac{\partial}{\partial x} (v \rho) dx \\ &= \int_{\alpha(t)}^{\beta(t)} \sigma \left( \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x} (v \rho) \right) dx. \end{aligned} \quad (6.25)$$

Substituting this into (6.23) we have that

$$\int_{\alpha(t)}^{\beta(t)} \left( \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(v\rho) \right) dx = 0, \quad (6.26)$$

and this is zero no matter what  $X_L$  and  $X_R$  we choose. Therefore, from the du Bois-Reymond lemma we conclude

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(v\rho) = 0. \quad (6.27)$$

This is the *continuity equation*, or the mass conservation equation, in spatial coordinates. It is also the traffic flow equation. As with traffic flow, the mathematical formulation is not complete because  $v$  is unknown. The difference now is that there are forces within the material, and by accounting for them we will be able to derive an equation for the velocity. This does not mean, however, that we are out of the constitutive law business. As will be seen shortly, that step has just been postponed until later in the development.

As a final comment, the Reynolds Transport Theorem was used to obtain (6.24), and this resulted in the evaluation of the integrand at the endpoints. In (6.25) these were then combined into a single integral using the Fundamental Theorem of Calculus. These two steps are the core of the integral method for deriving an equation of motion. It should be expected that any time the method is used that these steps will be present in the derivation.

### 6.4.1 Material Coordinates

There are situations when it is easier if the equations are expressed in material coordinates. To do this for the continuity equation let  $R(X, t)$  be the density in material coordinates. This means that  $R(X, t) = \rho(\mathcal{X}(X, t), t)$ , and from (6.10) we have that  $\frac{\partial R}{\partial t} = \rho_t + v\rho_x$ . Also, from Exercise 6.6, we have that  $v_x = \frac{\partial}{\partial t} \ln(1 + U_X)$ . With these two formulas, the continuity equation (6.27) transforms into

$$\frac{\partial R}{\partial t} + R \frac{\partial}{\partial t} \ln \left( 1 + \frac{\partial U}{\partial X} \right) = 0.$$

Solving this first-order equation for  $R$  yields

$$R(X, t) = \frac{R_0(X)}{1 + U_X}, \quad (6.28)$$

where  $R_0(X) = R(X, 0)$ . It would appear that by using material coordinates we have solved the continuity equation. This conclusion is correct although the

solution contains the displacement gradient and this is unknown until the momentum equation is solved.

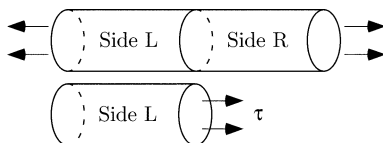
## 6.5 Momentum Equation

We will now introduce Newton's Second Law,  $F = ma$ , into the formulation. Actually, for our problem it is convenient to write this in momentum form as  $\frac{d}{dt}(mv) = F$ . To begin, we itemize the forces that are involved and how they usually enter the mathematical problem. They include:

*External Surface Forces.* These are, as the name implies, forces that affect the motion across the outer surface of the bar. For example, if you were to pick the bar up and stretch it you would be applying a surface force. Generally, for us these will only appear in the problem through boundary conditions.

*External Body Forces.* These affect all material points in the bar and will appear in the equation as a known forcing function. In the formulation below  $f(x, t)$  will represent the external body forces per unit mass, and so,  $\rho f$  is the resultant body force per unit volume. The standard example of an external body force is gravity, in which case  $f = -g$ .

*Internal Forces.* These are the forces that the constituents making up the bar exert on each other. For example, if the bar is stretched, then the material points in the bar pull on each other in an effort to restore the bar to its original length. To get a handle on these forces, note that given any cross-section, except those at the ends, you can separate the bar into a left and right side as shown in Fig. 6.3. In this context the left side can be thought of as being pulled (or pushed) by the material on the right. Although we do not yet know exactly what this force is, let  $\tau(x, t)$  denote its value per unit area. Because  $\tau$  has the dimensions of force/area it is a *stress* function. At the moment it is unknown and in the next section we will discuss at some length how to remedy this situation. Nevertheless, a few things can be said about the stress. Newton's Third Law states that for every action, there is an equal and opposite reaction. What this means, in regard to Fig. 6.3, is that if the stress on  $L$  due to  $R$  is  $\tau$ , then  $-\tau$  is the stress of  $L$  on  $R$ . The convention is that a force in the positive  $x$ -direction is positive. Consequently, if  $\tau > 0$ , then  $R$  is pulling on  $L$ , while if  $\tau < 0$ , then  $R$  is pushing on  $L$ .



**Fig. 6.3** Any internal material cross-section can be thought of as separating the bar into a left (L) and right (R) side. As shown, the bar is being stretched and this results in a stress  $\tau$  on side L due to the material on side R

The assumption made here is that the time rate of change of the momentum equals the sum of forces on the material. To understand the mathematical consequences of this assumption suppose that at  $t = 0$  we identify a segment of the bar, and assume this is an interval of the form  $X_L \leq x \leq X_R$ . At any later time this segment occupies an interval  $\alpha(t) \leq x \leq \beta(t)$ , where the endpoints are determined from the formulas  $\alpha(t) = \mathcal{X}(X_L, t)$  and  $\beta(t) = \mathcal{X}(X_R, t)$ . The total momentum of this segment is

$$\int_{\alpha(t)}^{\beta(t)} \sigma \rho v dx, \quad (6.29)$$

and the total body force on the segment is

$$\int_{\alpha(t)}^{\beta(t)} \sigma \rho f dx. \quad (6.30)$$

The force on the left end of the segment is  $-\sigma \tau(\alpha, t)$ , and on the right end it is  $\sigma \tau(\beta, t)$ . Therefore, from Newton's Second Law we obtain the equation

$$\frac{d}{dt} \int_{\alpha(t)}^{\beta(t)} \sigma \rho v dx = \int_{\alpha(t)}^{\beta(t)} \sigma \rho f dx + \sigma \tau(\beta, t) - \sigma \tau(\alpha, t). \quad (6.31)$$

Using the same steps as in (6.25), differentiation of the integral on the left-hand side of the above equation yields

$$\frac{d}{dt} \int_{\alpha(t)}^{\beta(t)} \sigma \rho v dx = \int_{\alpha(t)}^{\beta(t)} \sigma \left( \frac{\partial(\rho v)}{\partial t} + \frac{\partial}{\partial x}(v^2 \rho) \right) dx. \quad (6.32)$$

As for the right-hand side of (6.31), we can write it as

$$\int_{\alpha(t)}^{\beta(t)} \sigma \rho f dx + \sigma \tau(\beta, t) - \sigma \tau(\alpha, t) = \int_{\alpha(t)}^{\beta(t)} \sigma \left( \rho f + \frac{\partial \tau}{\partial x} \right) dx. \quad (6.33)$$

Combining (6.32) and (6.33) we have that

$$\int_{\alpha(t)}^{\beta(t)} \sigma \left( \frac{\partial(\rho v)}{\partial t} + \frac{\partial}{\partial x}(v^2 \rho) - \rho f - \frac{\partial \tau}{\partial x} \right) dx = 0. \quad (6.34)$$

This holds for any segment, and so from the du Bois-Reymond lemma it follows that

$$\frac{\partial(\rho v)}{\partial t} + \frac{\partial}{\partial x}(v^2 \rho) - \rho f - \frac{\partial \tau}{\partial x} = 0. \quad (6.35)$$

Using the continuity equation (6.27), this can be written as

$$\rho \frac{Dv}{Dt} = \rho f + \frac{\partial \tau}{\partial x}, \quad (6.36)$$

where the material derivative  $\frac{Dv}{Dt}$  is given in (6.11). This is the momentum equation for the bar expressed in spatial coordinates.

### 6.5.1 Material Coordinates

It is straightforward to rewrite the momentum equation in material coordinates. We know that  $\rho(x, t) = R(X, t)$  and  $\frac{Dv}{Dt} = \frac{\partial V}{\partial t}$ . As for the stress, let  $T(X, t)$  be the stress in material coordinates. It is assumed that the value of the force illustrated in Fig. 6.3 between  $L$  and  $R$  is the same for the material and spatial systems. Consequently, since the cross-sectional area is constant, it follows that  $T(X, t) = \tau(x, t)$  for  $x = \mathcal{X}(X, t)$ . With this, and Table 6.1,

$$\frac{\partial \tau}{\partial x} = \frac{1}{1 + U_X} \frac{\partial T}{\partial X}. \quad (6.37)$$

Similarly, letting  $F(X, t)$  be the body force in material coordinates, then  $f(x, t) = F(X, t)$ . Substituting all of this information into (6.36), and making use of the continuity equation (6.38), one obtains

$$R_0 \frac{\partial^2 U}{\partial t^2} = R_0 F + \frac{\partial T}{\partial X}, \quad (6.38)$$

where  $R_0(X)$  is the initial density.

## 6.6 Summary of the Equations of Motion

To summarize the formulation of the equations of motion up to this point, we have found that in spatial coordinates the continuity and momentum equations are, respectively,

**Spatial Version:**

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(v\rho) = 0 \quad (6.39)$$

$$\rho \left( \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} \right) = \rho f + \frac{\partial \tau}{\partial x} \quad (6.40)$$

**Table 6.2** Variables appearing in the equations of motion

Variable	Material	Spatial
Coordinates	$X, t$	$x, t$
Displacement	$U(X, t)$	$u(x, t)$
Velocity	$V(X, t)$	$v(x, t)$
Density	$R(X, t)$	$\rho(x, t)$
Stress	$T(X, t)$	$\tau(x, t)$
External body force	$F(X, t)$	$f(x, t)$

### Material Version:

$$R(X, t) = \frac{R_0}{1 + U_X} \quad (6.41)$$

$$R_0 \frac{\partial^2 U}{\partial t^2} = R_0 F + \frac{\partial T}{\partial X} \quad (6.42)$$

The variables in these equations are defined in Table 6.2. Also,  $R_0 = R(X, 0)$ .

Comparing the spatial and material versions it is easy to come to the conclusion that the material form is easier to use. For one reason the continuity equation does not need to be solved. Also, the material version of the momentum equation does not explicitly contain nonlinear terms, such as  $vv_x$  appearing in the spatial momentum equation. This does not mean, however, that the material version is linear as we have yet to determine how the stress is related to the density and displacement. Even so, the evidence appears to support the conclusion that the material version is easier to use. The fact is, however, that there are situations where the spatial version is preferred. Examples are easy to find in fluid dynamics because it is common when studying fluid motion to observe the flow from a fixed spatial position. In such cases the spatial version is more natural. At this point we will keep an open mind on the subject and use whichever seems to produce the easiest problem to solve.

A comment is necessary about the earlier statement that  $T(X, t) = \tau(x, t)$ . Both are stresses, which mean that they are measures of force per area. How they differ is that  $T(X, t)$  is based on the original area, and  $\tau$  uses the current area. In the derivation of (6.42),  $\sigma$  is assumed constant, so the original and current areas are the same. If  $\sigma$  depends on  $x$ , then, as shown in Exercise 6.9, the momentum equation changes because the original area  $\sigma(X)$  and the current area  $\sigma(\mathcal{X})$  differ. The formula relating these two stress in the case of three-dimensional motion is given in (8.88).

As another observation, the mathematical problem consists of two equations involving several variables. The body force term in the momentum equation is assumed known. This leaves us with what looks to be three dependent variables to solve for: the density, the velocity, and the stress. So, we have either one too many unknowns or we are short one equation. The approach taken in continuum mechanics is to introduce a constitutive law for the stress, which relates it with the other two dependent variables. This is not a new situation for us as we had to do something similar in the traffic flow problem. Before doing this, we examine the steady-state solution.

## 6.7 Steady-State Solution

A simple yet informative problem involves the steady state. This is the situation where the bar has come to rest, so the variables are independent of time. Assuming there are no body forces, then the momentum equation (6.42) reduces to

$$\frac{\partial T}{\partial X} = 0. \quad (6.43)$$

Therefore, at a steady state with no body forces the stress  $T$  is a constant. To determine the value of  $T$  we need to know what was done to the bar to cause it to deform. In other words, we need to know the boundary conditions. Experimentally there are two commonly used testing methods, and they are considered in the examples below.

*Example (Prescribed Displacement)* Consider the situation of when a bar of length  $\ell_0$  is stretched to length  $\ell$  and held in this position. This is illustrated in Table 6.3. As shown, the bar initially occupies the interval  $0 \leq x \leq \ell_0$ , and it is then stretched, so it occupies  $0 \leq x \leq \ell$ . Because the left end of the bar is held fixed, it is relatively easy to write down the corresponding boundary condition. In material coordinates it is

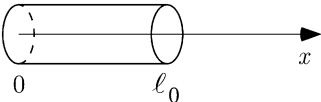
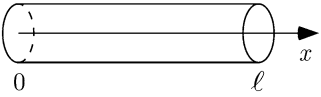
$$U|_{X=0} = 0. \quad (6.44)$$

The bar is stretched to length  $\ell$ , and for the steady-state problem this translates into the following boundary condition:

$$U|_{X=\ell_0} = \ell - \ell_0. \quad (6.45)$$

Although we know the displacement at the two ends we do not know how to relate this to the stress or displacement within the bar. This will have to wait until we specify the constitutive law for the stress. One last comment to make is that it is

**Table 6.3** The boundary conditions for a bar before and after being stretched, using material and spatial coordinates

Configuration	Time	Material	Spatial
	$t = 0$	$0 \leq X \leq \ell_0$	$0 \leq x \leq \ell_0$
		$U _{X=0} = 0$	$u _{x=0} = 0$
		$U _{X=\ell_0} = 0$	$u _{x=\ell_0} = 0$
	$t = \infty$	$0 \leq X \leq \ell_0$	$0 \leq x \leq \ell$
		$U _{X=0} = 0$	$u _{x=0} = 0$
		$U _{X=\ell_0} = \ell - \ell_0$	$u _{x=\ell} = \ell - \ell_0$



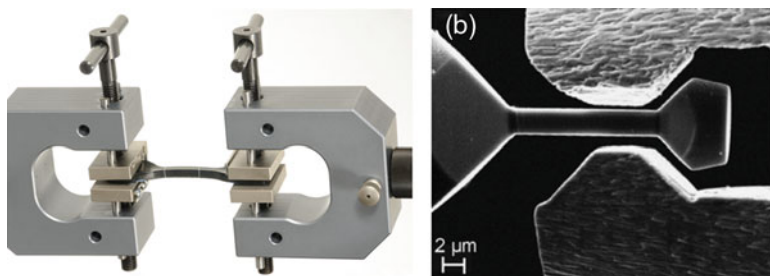
possible to express this steady-state problem in terms of the spatial variables, and the associated boundary conditions are given in Table 6.3. ■

*Example (Prescribed Stress)* Another common method of stretching, or compressing, a bar is to apply a force on one of its ends. A situation where this arises is when the bar is vertical and a weight is attached to the lower end, which causes the bar to stretch. To put this into mathematical terms, assume the bar is held at  $x = 0$ , so the condition at this end is the same as before; namely,  $U(0, t) = 0$ . At the other end assume a constant force  $F_0$  is applied. The boundary condition in this case is  $T(\ell_0, t) = F_0/\sigma$ . At steady state the stress is constant throughout the bar, and this means  $T(X, t) = F_0/\sigma$ . Because we have been able to determine the stress in the bar we have gotten a bit further in solving the problem than in the stress relaxation example. However, we still do not know the displacement of the bar except at  $x = 0$ . Again, the issue is how to relate the stress to the displacement, and this is the reason for needing a constitutive law. ■

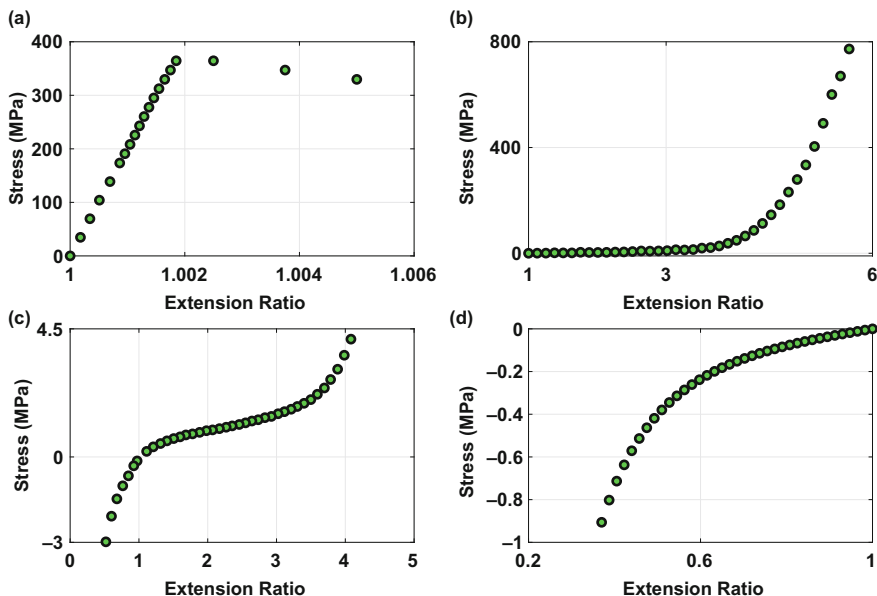
## 6.8 Constitutive Law for an Elastic Material

To complete the formulation we need to determine a constitutive law for the stress. This step requires more thought than is usually realized. It is not uncommon in modeling textbooks simply to state a law and then get on with the mathematics. Such an approach ignores some of the more interesting, and important, questions that arise in applied mathematics. The reason is that determining a constitutive law requires close interaction of the mathematics with experiments, and even after decades of research the principles that underpin constitutive laws are still not completely understood.

The first question to address is what properties of the solution we can determine based on what is known about the problem so far. The objective is to compare the



**Fig. 6.4** Determining the stress-strain curve experimentally involves using a computer controlled gripping mechanism to either stretch or compress the material. The scale over which this is done is enormous, as illustrated by the steel sample on the scale of centimeters, on the left, and copper crystal sample on the scale of microns, on the right (Kiener et al., 2008)



**Fig. 6.5** Stress measured as a function of the extension ratio (6.46), for (a) steel, (b) capture silk from a spider web (Blackledge and Hayashi 2006), (c) rubber (Raos 1993), and (d) articular cartilage (Kwan 1985)

mathematical model with what is determined experimentally. An obvious choice is the steady-state solution obtained from stretching, or compressing, a bar from length  $\ell_0$  to length  $\ell$ . As shown in the preceding section, at steady state the stress  $T$  is constant throughout the bar. This is useful information because one of the most common material testing experiments involves measuring the steady-state stress as a function of  $\ell$ . Examples of the gripping mechanisms that are used, and the materials tested, are shown in Fig. 6.4. For testing in tension, so  $\ell > \ell_0$ , samples are cut into small strips and inserted into a computer-controlled testing machine. The range of such experimental systems is enormous, from huge machines that are capable of testing samples the size of a car, down to microscopic systems that are used to test single molecules. Given this, it is not surprising that the types of materials tested in this way are quite varied, and four examples are given in Fig. 6.5. In this figure the measured stress is given as a function of the *extension ratio*

$$\lambda = \frac{\ell}{\ell_0}. \quad (6.46)$$

With this, the material is in tension if  $\lambda > 1$ , and it is in compression if  $\lambda < 1$ . The range of extension in the figure varies significantly with the material. For example, the range for steel is much smaller than it is for rubber. This difference is not surprising. Also, steel has the odd feature that the stress starts decreasing at larger

extension ratios. This is due to the metal being pulled apart, and it is characteristic of what are called ductile materials. In contrast, brittle materials, such as glass, simply break. We will assume the displacements are not so extreme as to cause this irreversible behavior to occur, and when the force is removed that the material will return to its original shape.

In looking at the data in Fig. 6.5 it is apparent that, for the materials shown, the stress increases with the imposed displacement  $U = \ell - \ell_0$ . This is consistent with the everyday observation that the more you stretch something the greater it resists. Based on this it might seem reasonable that for our constitutive law we should assume  $T = T(U)$ . The fact is, however, that this is not possible. We found earlier that at a steady state the stress is constant. If  $T = T(U)$ , then this would require that the displacement is also constant. The difficulty with this is that we require  $U = 0$  at the left end of the bar and  $U = \ell - \ell_0$  at the right end. It is therefore impossible for the displacement to be constant and, consequently, it is not possible to assume the stress is a function only of the displacement.

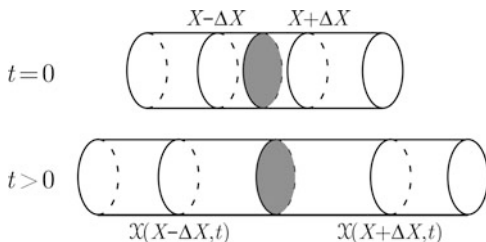
### 6.8.1 Derivation of Strain

A way to correct the difficulty discussed in the previous paragraph is to assume the stress depends on the relative displacement. There are various ways to measure relative displacement and an example is  $(\ell - \ell_0)/\ell_0$ , which compares the displacement  $\ell - \ell_0$  to the original length  $\ell_0$ . There are other ratios for measuring relative displacement and some of the more commonly used are listed in Table 6.4. At this point there is no clear reason why you would want to pick one over another and we will use the Lagrangian strain, leaving the others for the exercises. For cultural reasons it is worth saying something about the names given the different strain measures. The ratio used to derive the Lagrangian strain is known in the literature as the engineering or nominal strain. You will also see the Hencky strain referred to as the natural, or true, strain. In this text whenever referring to strain it is understood we are using the Lagrangian strain as defined in Table 6.4.

**Table 6.4** Various strain measures used in continuum mechanics

Name	Ratio	Definition
Lagrangian strain	$(\ell - \ell_0)/\ell_0$	$\epsilon = U_X$
Eulerian strain	$(\ell - \ell_0)/\ell$	$\epsilon_e = u_x$
Green strain	$(\ell^2 - \ell_0^2)/(2\ell_0^2)$	$\epsilon_g = U_X + \frac{1}{2}U_X^2$
Almansi strain	$(\ell^2 - \ell_0^2)/(2\ell^2)$	$\epsilon_a = u_x - \frac{1}{2}u_x^2$
Midpoint strain	$2(\ell - \ell_0)/(\ell + \ell_0)$	$\epsilon_m = U_X/(1 + \frac{1}{2}U_X)$
Hencky strain	$\ln(\ell/\ell_0)$	$\epsilon_h = \ln(1 + U_X)$

**Fig. 6.6** A segment starts off centered at  $X$  with length  $2\Delta X$ . At time  $t$  the segment has length  $\mathcal{X}(X + \Delta X, t) - \mathcal{X}(X - \Delta X, t)$



The basic assumption for our constitutive law is that the stress depends on the relative displacement  $(\ell - \ell_0)/\ell_0$ . To be more precise, we will assume that the stress at a material point depends on the relative displacement in the immediate vicinity of this point. To translate this into mathematical terms, given a cross-section located initially at  $X$ , consider cross-sections at  $X \pm \Delta X$  that are just to the left and right (see Fig. 6.6). After time  $t$ , the cross-section on the right moves to  $\mathcal{X}(X + \Delta X, t)$ , and the one on the left moves to  $\mathcal{X}(X - \Delta X, t)$ . The length of this small segment of the bar is  $\mathcal{X}(X + \Delta X, t) - \mathcal{X}(X - \Delta X, t)$ , while the original length was  $2\Delta X$ . Recalling that  $\mathcal{X} = X + U$ , then the ratio for the relative displacement is

$$\begin{aligned} \frac{\text{new length} - \text{original length}}{\text{original length}} &= \frac{\mathcal{X}(X + \Delta X, t) - \mathcal{X}(X - \Delta X, t) - 2\Delta X}{2\Delta X} \\ &= \frac{U(X + \Delta X, t) - U(X - \Delta X, t)}{2\Delta X}. \end{aligned} \quad (6.47)$$

Assuming  $\Delta X$  is small, and using Taylor's theorem,

$$\begin{aligned} U(X \pm \Delta X, t) &= U(X, t) \pm \Delta X \frac{\partial U}{\partial X}(X, t) + \frac{1}{2} \Delta X^2 \frac{\partial^2 U}{\partial X^2}(X, t) \pm \frac{1}{6} \Delta X^3 \frac{\partial^3 U}{\partial X^3}(X, t) + \dots \end{aligned}$$

Introducing these into (6.47) we obtain

$$\frac{\text{new length} - \text{original length}}{\text{original length}} = \frac{\partial U}{\partial X}(X, t) + O(\Delta X^2). \quad (6.48)$$

Therefore, a local measure of the relative distortion in the vicinity of a material point is

$$\epsilon = \frac{\partial U}{\partial X}, \quad (6.49)$$

which is known as the Lagrangian strain, or in this textbook, simply the strain.

With the definition of strain given in (6.49), the assumed constitutive law for the stress is  $T = T(\epsilon)$ . A material for which this holds is said to be *elastic*. The momentum equation, in material coordinates, for an elastic material is

$$R_0 \frac{\partial^2 U}{\partial t^2} = R_0 F + T' \left( \frac{\partial U}{\partial X} \right) \frac{\partial^2 U}{\partial X^2}. \quad (6.50)$$

This is a wave equation for  $U$ , and it is nonlinear if the stress is a nonlinear function of the strain. Not just any function can be used for the stress, and later in the chapter we will investigate some of the restrictions that must be imposed on how it depends on strain.

One last useful piece of information concerns the extension ratio (6.46). Given the result in (6.48), when deriving the continuum formulation by letting  $\Delta X \rightarrow 0$ , the extension ratio  $\lambda$  turns into  $1 + \epsilon$ . The reason for pointing this out is that in the simplification of the constitutive law for the stress that is given below, we will investigate how the measured stress behaves in the neighborhood of  $\lambda = 1$ . In the continuum formulation this is equivalent to looking at how the stress behaves around  $\epsilon = 0$ .

### 6.8.2 Material Linearity

With the assumption that  $T = T(\epsilon)$ , we return to the stress curves in Fig. 6.5. The dependence of  $T$  on  $\epsilon$  clearly depends on the material. This is reasonable as the morphological and mechanical characteristics of these materials are markedly different. Even so, there is a region for each material, containing  $\epsilon = 0$ , where the stress is approximately a linear function of strain. The constitutive law in this case reduces to

$$T = E \frac{\partial U}{\partial X}, \quad (6.51)$$

where  $E$  is known as Young's, or the elastic, modulus. A material that follows this law is said to be linearly elastic. The momentum equation (6.50) in this case reduces to

$$R_0 \frac{\partial^2 U}{\partial t^2} = R_0 F + E \frac{\partial^2 U}{\partial X^2}. \quad (6.52)$$

This is a linear wave equation for the displacement  $U$ .

In the parlance of continuum mechanics, (6.51) is an assumption of material linearity. It should be understood that this constitutive law is not based on a requirement of a small strain. The strains for which (6.51) is valid need not be small, and it is only required that (6.51) furnishes an accurate approximation of the stress

**Table 6.5** Young's modulus and density of various materials, illustrating the range of values these parameters can have

Material	Young's modulus (GPa)	Density (kg/m <sup>3</sup> )
Diamond	1000	3500
Stainless steel	200	8030
Glass	65	2600
White oak	12	770
Beeswax	0.2	960
Rubber	0.007	1200
Silica aerogel	0.001	100

over the given range of strains. For example, in Fig. 6.5, steel is linearly elastic for strains up to about 0.002 while the capture fibers from a spider web are linear up to strains of approximately 2.0.

In terms of dimensional units, because strain is dimensionless, the elastic modulus has the same dimensions as stress. The basic unit for stress is the Pascal (Pa) and  $1 \text{ Pa} = 1 \text{ N/m}^2$ . Relative to the strength of most materials, however, a Newton (N) is relatively small. To illustrate this, 1 N is the force on an object with the mass of approximately an apple, and it takes a lot of apples to deform steel or glass a noticeable amount. For this reason, the elastic moduli of most materials run in the MPa or GPa range, where  $\text{M} = 10^6$  and  $\text{G} = 10^9$ . A few examples are given in Table 6.5. An observation coming from this table is that one should not assume that denser materials have larger elastic moduli. However, the density and modulus are connected through the molecular structure of the material, and this will be discussed later in the chapter.

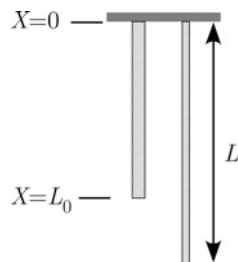
*Example (Steady-State Solution)* In Sect. 6.7 we were unable to solve the steady-state problem. The situation has improved with the introduction of the linear constitutive law in (6.51). The reduced momentum equation (6.47) now takes the form

$$\frac{\partial^2 U}{\partial X^2} = 0.$$

The solution of this that satisfies the boundary conditions (6.44) and (6.48) is  $U = (\ell - \ell_0)X/\ell_0$ . With this the stress is  $T = E(\ell - \ell_0)/\ell_0$ . Also, from (6.41), the density is  $R = ER_0/(T + E)$ . Therefore, as advertised, with the inclusion of the constitutive law, we have been able to determine the stress, displacement, and density in the bar. ■

*Example (Bungie Cord)* Apparently, for bungie jumpers it is more entertaining if the cord is long enough that the jumper comes very close to hitting the ground. Getting the right length cord is not a simple matter of just knowing the weight of the jumper and the height of the jump. The reason is that the weight of the cord will cause it to extend. To determine this, suppose the cord is hung with the upper end attached and the lower end free (see Fig. 6.7). Assume gravity is the only force present, and the

**Fig. 6.7** A bungee cord, originally with length  $L_0$ , stretches to length  $L$  after having been hung



cord starts off with a constant density  $R_0$  and length  $L_0$ . Also, assume that the cord is at rest and its length after being hung is  $L$ . Given that the cord is at a steady state, then the material momentum equation (6.42) takes the form

$$\frac{dT}{dX} = -R_0 g.$$

As for the boundary condition, the lower end is free and this means the stress is zero there. The boundary condition in this case is  $T = 0$  at  $X = L_0$ . From the above momentum equation, and the given boundary condition, we obtain

$$T = R_0 g (L_0 - X). \quad (6.53)$$

To determine the displacement of the cord we need to specify the constitutive law for the stress, and we will use (6.51). With this, and (6.53), we have that

$$\frac{dU}{dX} = \frac{R_0 g}{E} (L_0 - X). \quad (6.54)$$

Integrating this equation, and using that fact that the upper end is fixed, so  $U = 0$  at  $X = 0$ , then

$$U = \frac{R_0 g}{2E} X (2L_0 - X). \quad (6.55)$$

The bungee cord problem is now solved, and with the solution it is possible to determine just how far the cord will stretch. The displacement of the free end is obtained from (6.55) by taking  $X = L_0$ . The total length  $L$  of the stretched cord is obtained by adding this displacement to the original length  $L_0$ , and the result is

$$L = L_0 \left( 1 + \frac{R_0 g}{2E} L_0 \right). \quad (6.56)$$

This shows that a less stiff cord ( $E$  small) stretches longer. This is expected, but the above result shows that the stretched length is not a simple multiple of the modulus. For example, it does not happen that reducing the modulus by a factor of two causes

the length to double. Also, because of the nonlinear dependence of  $L$  on  $L_0$ , we have found that longer cords stretch proportionally longer than shorter cords. ■

### 6.8.3 Material Nonlinearity

It is of interest to consider the mathematical problem that results when a more physically realistic constitutive law is used. One possibility is to take a cubic of the form

$$T = E\epsilon + K\epsilon^3, \quad (6.57)$$

where  $E$  and  $K$  are positive constants. This curve has a passing resemblance to the stress curve for rubber shown in Fig. 6.5. In this case, the momentum equation (6.50) yields

$$R_0 \frac{\partial^2 U}{\partial t^2} = R_0 F + \left( E + K \left( \frac{\partial U}{\partial X} \right)^2 \right) \frac{\partial^2 U}{\partial X^2}. \quad (6.58)$$

Not unexpectedly, we now have a nonlinear wave equation for the displacement  $U$ .

Given typical boundary and initial conditions, the only way to solve (6.58) is numerically. This is easy to do, but this approach is limited to the specific problem that is solved. Often times it is of more interest to have a qualitative understanding of how the nonlinearity affects the motion. As an example, suppose there is no body force and  $R_0$  is constant. As will be shown in Sect. 7.1.1, the linear equation (so,  $K = 0$ ) can have traveling wave solutions of the form  $f(X - ct)$ . An often asked question is whether the nonlinear equation does as well. Answering this means you simply assume that the solution has the form of a traveling wave, and then see if it can satisfy the nonlinear wave equation. As a case in point, assuming  $U = f(X - ct)$ , one finds from (6.58) that  $f(\eta) = \alpha + \beta\eta$ . This is a solution but it is not a traveling wave similar to what is obtained in the linear problem. If you are convinced there are wave-like solutions, then you will need to be more clever in formulating an assumption that demonstrates this. Even though this approach has not been successful when applied to (6.58), it has been very useful for other problems involving nonlinear traveling waves. Those interested in this should consult Atkinson et al. (1981), Chiron (2012), or Griffiths and Schiesser (2009).

Another common approach used to help understand the effects of the nonlinearity is to assume it is weak. This has been used on numerous occasions in this textbook, examples being given in Sects. 1.5.2, 2.2.3, and 2.7. It is also used in the next example.

*Example (Nonlinear Bungee Cord)* In the bungee cord example considered in the last section, suppose it is assumed that the material is nonlinear and the constitutive law for the stress is given in (6.57). This does not change (6.53), but (6.54) now becomes



$$E \frac{dU}{dX} + K \left( \frac{dU}{dX} \right)^3 = R_0 g (L_0 - X). \quad (6.59)$$

This is an odd looking first-order nonlinear differential equation for  $U$ . Rather than attempt to solve it, we will make the assumption that the nonlinearity is weak. More specifically, it is assumed that  $|U_X| \ll 1$  and the displacement has an expansion of the form  $U \sim U_0 + U_1 + \dots$ , where  $U_1 \ll U_0$  (see Exercise 2.34 for a more rigorous derivation of the expansion). The first-order equation coming from (6.59) is just (6.54) but with  $U_0$  instead of  $U$ . The second-order equation is then

$$E \frac{dU_1}{dX} + K \left( \frac{dU_0}{dX} \right)^3 = 0.$$

Solving this for  $U_1$ , and using the boundary condition  $U_1 = 0$  at  $X = 0$ , then

$$U_1 = \frac{K}{4R_0g} \left( \frac{R_0g}{E} \right)^4 \left[ (L_0 - X)^4 - L_0^4 \right].$$

With this,

$$L = L_0 \left[ 1 + \frac{R_0g}{2E} L_0 - \frac{K}{4R_0g} \left( \frac{R_0g}{E} \right)^4 L_0^3 \right]. \quad (6.60)$$

So, depending on the value of  $K$ , the nonlinearity reduces the amount the cord is stretched compared to the case of a linear material. This is not a surprise, as the nonlinear constitutive law (6.57) states that  $T$  increases relative to the linear law. This means the material becomes stiffer as it is stretched, and so it will not stretch as much as it does for the linear law. ■

### 6.8.4 End Notes

The basic ideas underlying linear elasticity were developed by Robert Hooke, and (6.51) is sometimes referred to as Hooke's law. Given this, it might seem odd that the one parameter that appears in the equations is named after a physician named Thomas Young. The reason for this is that Hooke's original statement that "as is the extension, so is the force" implies that the force is proportional to displacement. For springs this might be acceptable but as we saw earlier this assumption is inapplicable to elastic bars. It was Young who interpreted it correctly using strain.

The statement that the stress is a linear function of strain depends on the strain and coordinate used in the formulation. For example, using (6.16), the constitutive law (6.51) expressed in spatial coordinates is

$$\tau = E \frac{u_x}{1 - u_x}. \quad (6.61)$$

Consequently, an assumption of material linearity using one strain measure does not necessarily mean it is linear using another strain measure.

In the experiments used to produce the data in Fig. 6.5, the experimenter waits until the motion stops before measuring the stress. This means that the constitutive law for the stress is determined using the steady-state response. Even so, the linear constitutive law (6.51) is assumed to apply even when the bar is in motion. If the stress also depends on rate variables, such as  $v$  or  $v_x$ , our approach of using the steady state to determine the constitutive law would miss this completely. There are materials that depend strongly on rate variables and examples are water, jello, and silly putty. To determine the correct rate dependence requires dynamic tests and several are commonly used in material testing. Exactly how this is done will be explained when we study viscoelasticity in the next chapter.

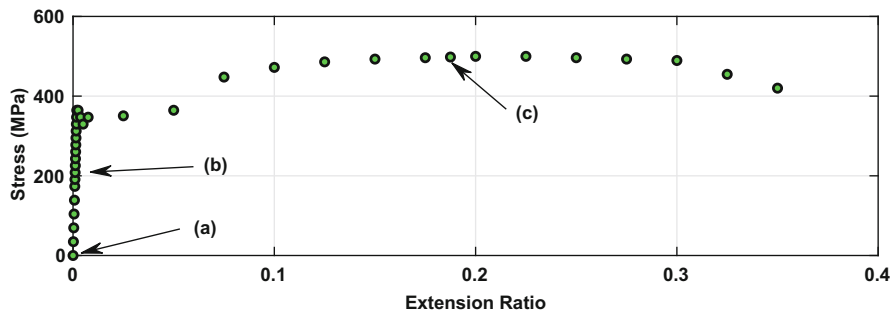
## 6.9 Morphological Basis for Deformation

The constitutive law used for the stress is a mathematical expression for how the material reacts to deformation. The materially linear assumption in (6.51) is routinely used to describe such diverse materials as steel, rubber, and skin. Given the differences in the atomic, or molecular, structure of these materials it is of interest to be able to understand how the substructural changes that take place during deformation give rise to the constitutive law for the stress.

### 6.9.1 Metals

Undoubtedly, the most studied metal is steel. A typical stress-strain curve for steel is given in Fig. 6.5a, but a more complete version is shown in Fig. 6.8. Because of the larger range of the extension ratio, the linear portion of the curve is not as evident as it is in Fig. 6.5a. However, what is apparent is that at larger strains the material is far from a simple linear function. It is also evident that the curve is not monotonic, and as explained in the next section this generates rather serious mathematical difficulties.

To understand how the microstructure of metal accounts for the observed deformation, the atoms in most metals are arranged in a periodic array, forming a lattice pattern. In this description the atoms are modeled as spheres. The dominant attractive force on the atoms is due to metallic bonding, which arises from the positively charged metal atoms sharing electrons. The resulting pairwise force has the form  $F_a = \alpha/r^{m+1}$ , where  $r$  is the separation distance between atoms, and  $\alpha$  is a constant determined by the electronic characteristics of the material. For many



**Fig. 6.8** The complete stress-strain curve for the steel sample shown in Fig. 6.5a. The atomic configuration at (a), (b), and (c) is shown in Fig. 6.10

metals  $m = 5$  or  $m = 6$ . There is also a repulsive force that comes into play if the electron shells of the atoms overlap, and it is based on the Pauli exclusion principle. The associated form of this force is  $F_r = -\beta/r^{n+1}$ , where  $\beta$  is a constant. The value of  $n$  depends on the material, and typical values are  $n = 10$  for copper and  $n = 11$  for silver (Qi et al. 1999). The resulting force is

$$F = \frac{\alpha}{r^{m+1}} - \frac{\beta}{r^{n+1}}, \quad (6.62)$$

where  $0 < m < n$ , and  $\alpha, \beta$  are positive constants.

In material science the properties of metals are often characterized using energy, and for this reason the force is written in terms of a potential function  $V$ . This is done by writing  $F = \frac{dV}{dr}$ , where

$$V = \frac{-\alpha}{mr^m} + \frac{\beta}{nr^n}. \quad (6.63)$$

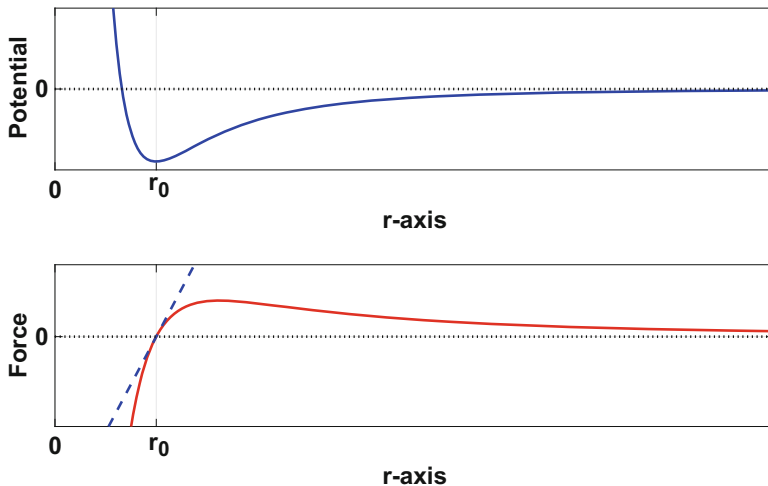
This function along with  $F$  are sketched in Fig. 6.9.

When no load is applied, so the atoms are in their equilibrium configuration, the two forces balance. Setting  $F = 0$  determines the equilibrium interatomic spacing  $r_0$ , and one finds that

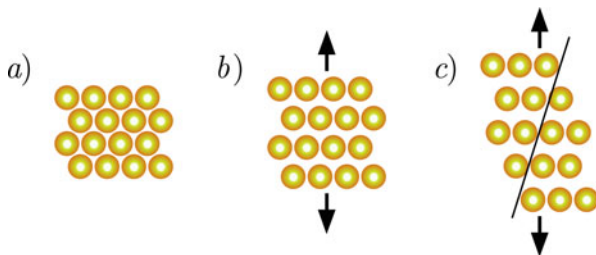
$$r_0 = \left( \frac{\beta}{\alpha} \right)^{\frac{1}{n-m}}. \quad (6.64)$$

This configuration is shown in Fig. 6.10a.

As the metal bar is stretched, the distance between the atomic layers increases, and the bonds between the atoms resist this change as described in (6.62). This is illustrated in Fig. 6.10b. If the load is not too large the bonds do not break, and when the load is removed the atoms return to their original positions in the lattice, shown in Fig. 6.10a. To relate the stress with the interatomic force, Fig. 6.10b shows



**Fig. 6.9** The force (6.62) and the potential (6.63) on the atoms in a metal. The dashed line is the tangent to the force curve at the point where  $F = 0$ . The slope of this line is used to obtain an approximation of the Young's modulus in (6.67)



**Fig. 6.10** The atomic configurations in a metal during deformation: (a) the atoms when no load is applied; (b) their position in the elastic region; and (c) the appearance of a slip plane for larger strain values

four cross-sections that are made up of atoms. To calculate the force between any two such cross-sections, note that there are approximately  $\sigma/(r_0^2)$  atoms in a square cross-section of area  $\sigma$ . So, the stress is approximately

$$T \approx \frac{\sigma}{r_0^2} \frac{F}{\sigma} = \frac{F}{r_0^2}. \quad (6.65)$$

In a similar manner, for  $r$  near  $r_0$ , the linear elastic law (6.51) can be approximated as

$$T(r_0) + (r - r_0)T'(r_0) \approx E \frac{r - r_0}{r_0}. \quad (6.66)$$

Combining (6.65) and (6.66), it follows that

$$\begin{aligned} E &\approx \left. \frac{1}{r_0} \frac{dF}{dr} \right|_{r=r_0} \\ &= \frac{(n-m)\alpha}{r_0^{m+3}}. \end{aligned} \quad (6.67)$$

Consequently, the elastic modulus has a strong dependence on the interatomic spacing  $r_0$ . Another observation is that the atomic mechanisms involved with tension, where  $r > r_0$ , are fundamentally different than those involved with compression, where  $r < r_0$ . This is why knowing the stress-strain function for a nonlinearly elastic material for tension provides little insight into what the stress function is for compression.

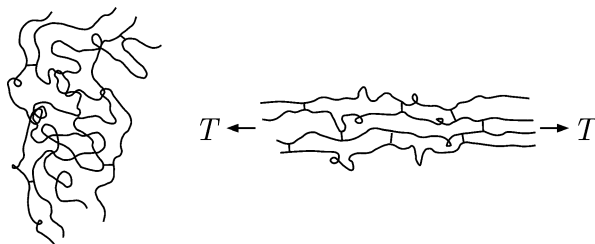
The largest load that, when removed, results in the atoms returning to their original configuration is known as the elastic limit. From Fig. 6.5a the elastic limit corresponds to an extension ratio of approximately 0.002. If a larger load is used slip planes will start to appear, and one is illustrated in Fig. 6.5c. As the name suggests, at a slip plane the atoms slide over each other along a plane. This is a permanent modification and if the load is removed the slip planes remain. In this situation the material is said to be plastic. Stretching the bar any farther produces more slip planes. Other defects in the atomic structure appear, including dislocations, and the specific events depend on the metal being tested. Eventually the metal is not capable of withstanding the stretching and breaks, a point material scientists call fracture. In Fig. 6.8 this happens when the extension ratio reaches about 0.35.

The function in (6.63) is a special case of what is known as a Sutton-Chen potential. We only considered what is effectively a nearest neighbor approximation using this potential, which means that we only considered the forces between a molecule and its nearest neighbor in the adjoining cross-section. A more realistic description would account for the other molecules, in which case  $F$  would consist of a sum of attractive and repulsive forces. It is worth noting that (6.63) includes well-known models, each accounting for different interatomic forces. One is the Lennard-Jones potential. This applies to an inert gas, where the attractive force is due to van der Waals bonding, where  $m = 6$  and  $n = 12$ . Another is the potential for ionic materials, such as NaCl, where  $m = 1$ . There has been considerable research in the last few years into what is called atomistic-based continuum theory, where the material's continuum properties are derived using interatomic potential functions. An introduction to this can be found in Giustino (2014) and Finnis (2010).

## 6.9.2 Elastomers

An elastomer is rubber made with a loosely cross-linked molecular structure. To explain what this means, natural rubber is made up of long individual molecules, or

**Fig. 6.11** Elastomer network, before and after the application of an axial load. The cross-links between the rubber molecules increase its ability to resist the load, and enable the network to return to its original configuration once the load is removed



more specifically, from long polymer chains. In effect, it is the molecular version of spaghetti. This changes if sulfur is added because this produces atomic bridges between the polymer chains. The consequence of this is a material that consists of long entangled molecules that are cross-linked, and a schematic of this is given in Fig. 6.11. Assuming that the number of cross-links is not too large, one produces what is known as an elastomer. Such materials are formed from a three-dimensional molecular network in which highly flexible molecules are connected at points provided by cross-links between the molecules.

In stretching an elastomer, the entangled polymer chains start to straighten. They are extendable but at large strains the cross-links limit the movement of the chains relative to one another. Consequently, upon the application and release of a stress, the molecules quickly revert to their normal crumpled form in the unstressed configuration, and this is the basis of the reversible high extensibility of elastomeric solids. This scenario applies to the materials in Fig. 6.5b, d. Both the capture silk and rubber offer relatively little resistance for extension ratios up to about 3. This is the interval over which the polymer molecules are uncoiled. Once that happens, and the cross-links become engaged, both materials show significant resistance and the stress increases almost exponentially. An example of a constitutive law incorporating this into the formulation is examined in Exercise 6.29. A more extensive investigation into the molecular contributions to the elastic behavior of an elastomer can be found in Mark and Erman (2007).

## 6.10 Restrictions on Constitutive Laws

One of the central problems in continuum modeling is finding the appropriate constitutive law for the stress. What is appropriate depends on what the model is describing. If the goal is to determine the deformation of a table due to the load of a computer, then the strains are likely so small that a linear theory can be used. On the other hand, if you are interested in the deflection of a trampoline, then the strains are likely so large that a nonlinear theory would be required. One question that arises in such cases is, what function should be used to describe this nonlinear behavior? As an example, for the data for rubber in Fig. 6.5c, the curve resembles a cubic. Based on this observation, one might assume that

$$T = a\epsilon + b\epsilon^3, \quad (6.68)$$

where  $\epsilon = \frac{\partial U}{\partial X}$ . On the other hand, the data for capture silk in Fig. 6.5b look to follow more of an exponential function, and a possible constitutive law that could be used in this case is

$$T = a(e^{b\epsilon} - 1). \quad (6.69)$$

One of the standard answers to the question of what function to use is that it is relatively simple, and it does a reasonable job describing the stress-strain data. Although reproducing the experimental results is a worthy goal, you want the model to also describe the motion in situations for which you do not have data. As an example, we know that strains must satisfy  $-1 < \epsilon < \infty$ . So, suppose one of the above nonlinear functions is used to fit data in the range  $-0.5 < \epsilon < 50$ . It is questionable that either one would successfully describe what happens for  $-1 < \epsilon < -0.5$  because both predict a finite stress when the material is compressed to zero (i.e., when  $\epsilon \rightarrow -1$ ). Because of this, it is worth imposing a requirement on the constitutive law to guarantee the right behavior under the extreme condition of letting  $\epsilon \rightarrow -1$ . It is the objective of this section to develop some general requirements that can be used to help formulate the constitutive law.

### 6.10.1 Frame-Indifference

Considering what happens to the stress when  $\epsilon \rightarrow -1$  falls into the extreme behavior category. Another category relates to the assumptions made in defining the stress. As stated in Sect. 6.5,  $T$  is due to the forces between molecules in the material as they move relative to each other (see Fig. 6.10). So, suppose it's assumed that  $T = T(U)$ . This would mean that if all of the points of a bar move a constant distance  $U$ , then the stress at each point would be  $T(U)$ . However, the relative positions of the points do not change in this case, so there should be no change in the stress. The implication is that  $T(U)$  is identically zero, which clearly is not possible (see Fig. 6.5). By considering a uniform velocity, one can use a similar argument to rule out a constitutive law of the form  $T = T(V)$ .

The question is, what variables can  $T$  depend on that are consistent with the definition of the stress. The usual approach to answer this relies on a mathematical formalism involving how the constitutive law changes with the observer. To explain, the argument used to rule out using  $U$  or  $V$  involved the observation that  $T$  should not change if the material undergoes a rigid body motion. In  $\mathbb{R}^3$ , rigid body motions involve rotations and translations, but in  $\mathbb{R}$  they are just translations. Consequently, the observers in our formulation are translations of each other. This gives us the following assumption.

**Principle of Material Frame-Indifference.** *The form of the constitutive law for the stress is independent of the observer.*

To translate this into a mathematical requirement, suppose two coordinate systems  $(x, t)$  and  $(x^*, t^*)$  are related through a change of coordinates given as  $x^* = x + b(t)$  and  $t^* = t - t_0$ . Using a superscript  $*$  to denote the value of a quantity in the  $(x^*, t^*)$  system, then a function  $f(x, t)$  is *frame-indifferent* if  $f(x, t) = f^*(x^*, t^*)$  for all  $t_0$  and permitted  $b(t)$ . The sticking point here is what “permitted” means. It is common in continuum mechanics to allow  $b(t)$  to be any smooth function, which means that the change of coordinates corresponds to what is called an *Euclidean transformation*. In contrast, the usual assumption made in Newtonian physics is that  $b(t)$  is limited to a linear function of time. In this case, the change of coordinates corresponds to a *Galilean transformation*. Which transformation to require has generated numerous “I’m right, and you have no idea what you are talking about” research papers. For us, given the relatively simple applications we are considering, either can be used without affecting the resulting conclusions.

To guarantee that the stress is frame-indifferent we will limit the constitutive law to only depend on frame-indifferent variables. This means we need a list of what is, and what is not, frame-indifferent.

*Example 1* The displacement function is not frame-indifferent. To explain why, in each coordinate system we have a different position function, so  $x = \mathcal{X}(X, t)$  and  $x^* = \mathcal{X}^*(X, t^*)$ . Given that the change of coordinates is  $x^* = x + b(t)$ , then the position functions satisfy  $\mathcal{X}^* = \mathcal{X} + b$ . Expressing this equation in terms of the displacement function we have that

$$U^*(X, t^*) = U(X, t) + b(t). \quad (6.70)$$

Given (6.4), in terms of spatial coordinates, the above equation takes the form  $u^*(x^*, t^*) = u(x, t) + b(t)$ . To be frame-indifferent we must be able to conclude that  $u^* = u$ , no matter what  $b$  we select. Clearly this does not happen and the conclusion is that the displacement is not frame-indifferent. ■

*Example 2* The velocity function is not frame-indifferent. This follows by taking the time derivative of (6.70) and concluding that  $V^* = V + b'(t)$ . From (6.5), this can be written as  $v^* = v + b'(t)$ . Given that  $b'(t)$  is not necessarily zero, it follows that the velocity is not frame-indifferent. ■

*Example 3* The strain function  $\frac{\partial U}{\partial X}$  is frame-indifferent. This follows by taking the  $X$  derivative of (6.70) and concluding that  $U_X^* = U_X$ . Since, from Table 6.1,

$$\frac{\partial u}{\partial x} = \frac{U_X}{1 + U_X}$$



we conclude that the strain functions  $\frac{\partial u}{\partial x}$  and  $\frac{\partial U}{\partial X}$  are frame-indifferent. Therefore, the assumption underlying the constitutive law for an elastic material satisfies the Principle of Material Frame-Indifference. ■

*Example 4* The function  $\frac{\partial \rho}{\partial t}$  is not frame-indifferent. First note that the density is assumed to be frame-indifferent. The reason is that mass and volume are invariant under rigid body motions. So,  $\rho(x, t) = \rho^*(x^*, t^*) = \rho^*(x + b, t - t_0)$ . Consequently,

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= \frac{\partial \rho^*}{\partial t} + \frac{\partial \rho^*}{\partial x^*} \frac{\partial x^*}{\partial t} \\ &= \frac{\partial \rho^*}{\partial t^*} + b'(t) \frac{\partial \rho^*}{\partial x^*}. \end{aligned}$$

Any change of coordinates with  $b' \neq 0$  means  $\rho_t \neq \rho_{t^*}^*$ , and so this function is not frame-indifferent. ■

Given this requirement on the constitutive law it is worth having a small list of functions that are frame-indifferent. Functions that are frame-indifferent include

$$\rho, \frac{D\rho}{Dt}, \frac{\partial U}{\partial X}, \frac{\partial u}{\partial x}, \frac{\partial V}{\partial X}, \frac{\partial v}{\partial x}. \quad (6.71)$$

Functions that are not frame-indifferent include

$$\frac{\partial \rho}{\partial t}, U, u, V, v. \quad (6.72)$$

Are there materials that use multiple frame-indifferent functions in the constitutive model? The answer is yes, and they are very common. A simple example is a viscoelastic material, where one assumes the stress depends on the strain  $U_X$  and the strain rate  $V_X$ . Examples such as this are explored in the next chapter.

### 6.10.2 Entropy Inequality

There are several other principles used to formulate constitutive laws. We will only consider one more, and it is related to the second law of thermodynamics. This requires the introduction of three new variables, and the first is connected with the energy. As with all mechanical systems, the energy involves both kinetic and potential components. It is relatively easy to identify the kinetic energy density, and it is  $\frac{1}{2}\rho v^2$ . The potential energy has multiple sources, and one comes from the external forcing. Another comes from the ability of the material to store energy, in the same way a spring stores energy when it is compressed. Because this component arises from the properties of the material, it is known as the internal energy. We want

to determine this in our continuum theory, and with this in mind let  $\chi(x, t)$  be the *internal energy density* per unit mass.

Like the density and momentum, the energy is assumed to satisfy a balance law, and it is

$$\frac{d}{dt} \int_{\alpha(t)}^{\beta(t)} \sigma \rho \left( \frac{1}{2} v^2 + \chi \right) dx = \int_{\alpha(t)}^{\beta(t)} \sigma \rho v f dx + \sigma v \tau \Big|_{x=\alpha}^{x=\beta}. \quad (6.73)$$

In words, the above equation states that the rate of change of the total energy of a material segment equals the sum of the rate of work of the external forces and the rate of work done by the forcing on the ends of the segment. Using the same argument employed to derive the continuity and momentum equations, the above expression results in the following equation:

$$\rho \frac{D\chi}{Dt} = \tau \frac{\partial v}{\partial x}. \quad (6.74)$$

This gives us an equation that can be solved to determine the function  $\chi$ .

The second variable that needs to be introduced is  $\eta(x, t)$ , which is the *entropy density* per unit mass. As expressed in the second law of thermodynamics, it is assumed that the entropy does not decrease. In other words, it is assumed that

$$\frac{D\eta}{Dt} \geq 0. \quad (6.75)$$

In continuum mechanics this is known as the Clausius-Duhem inequality. It is assumed here that there is no supply or flux of entropy. This can occur, for example, when there is heat flow in the system. In our development, the thermal effects are omitted.

The third, and final, function that needs to be introduced is the *Helmholtz free energy density*  $\psi$ , defined as

$$\psi = \chi - \theta \eta, \quad (6.76)$$

where  $\theta$  is the absolute temperature. Consistent with our earlier assumptions,  $\theta$  is assumed to be constant. The reason for calling  $\psi$  the free energy is that it represents the energy remaining to do work after accounting for what is invested in the entropic state of the material (Table 6.6).

**Table 6.6** Variables used in the formulation of the reduced entropy inequality

Variable	Material	Spatial
Internal energy density	$\mathcal{U}(X, t)$	$\chi(x, t)$
Entropy density	$\mathcal{N}(X, t)$	$\eta(x, t)$
Absolute temperature	$\Theta$	$\theta$
Helmholtz free energy	$\Psi$	$\psi$

Solving (6.76) for  $\eta$ , and then substituting the result into the Clausius-Duhem inequality (6.75), yields

$$-\rho \frac{D\psi}{Dt} + \tau \frac{\partial v}{\partial x} \geq 0. \quad (6.77)$$

This is known as the *reduced entropy inequality*. In material coordinates this inequality takes the form

$$-R_0 \frac{\partial \Psi}{\partial t} + T \frac{\partial V}{\partial X} \geq 0, \quad (6.78)$$

where  $\Psi$  is the material form of the Helmholtz free energy. It is assumed here that the spatial and material forms of the free energy functions give the same value. Therefore, if a cross-section that starts at  $X$  is currently located at  $x = \mathcal{X}(X, t)$ , then  $\Psi(X, t) = \psi(x, t)$ .

We are now in position to state the second requirement imposed on constitutive laws.

**Principle of Dissipation.** *A constitutive law must satisfy the reduced entropy inequality (6.77), or equivalently (6.78), for all physically consistent values of its arguments.*

As an example of physical consistency, the density values are required to be nonnegative. An illustration of when this comes up is given below, when obtaining (6.83) from (6.82).

Now comes the question of exactly how we use this condition because it involves the yet to be determined Helmholtz free energy  $\psi$ . We will show that in certain cases the stress  $\tau$  can be determined from  $\psi$ . This means that instead of formulating a constitutive law for the stress, we can specify one for  $\psi$ , and then use this to determine  $\tau$ . In doing this it is assumed that the constitutive law for  $\psi$  depends on the same variables used for the stress.

*Example* For an elastic material, the general form of the constitutive law in spatial coordinates is  $\tau = \tau(u_x)$ . The corresponding assumption for the constitutive law for the free energy function is  $\psi = \psi(u_x)$ . To see what this gives us, note, using the chain rule and Exercise 6.6(f),

$$\begin{aligned} \frac{D\psi}{Dt} &= \psi' \frac{D}{Dt} u_x \\ &= \psi' (1 - u_x) v_x. \end{aligned}$$

With this, (6.77) reduces to

$$[\tau - \rho(1 - u_x)\psi'(u_x)]v_x \geq 0. \quad (6.79)$$

According to the Principle of Dissipation, this inequality must hold for all values of  $v_x$ . For example, it must hold when  $v_x = 1$ , and when  $v_x = -1$ . Because the quantity in the square brackets does not depend explicitly on  $v_x$ , it must be that  $\tau - \rho(1 - u_x)\psi'(u_x) = 0$ . Therefore, for an elastic material, the stress is determined from the free energy as follows:

$$\tau = \rho(1 - u_x)\psi'(u_x). \quad \blacksquare \quad (6.80)$$

In the above example spatial coordinates were used. If one uses material coordinates, and assumes that the material is elastic, then the constitutive law has the form  $T = T(\epsilon)$ . The corresponding assumption for the free energy function is  $\Psi = \Psi(\epsilon)$ . Using (6.78), and an argument similar to the one used in the above example, one finds that

$$T = R_0\Psi'(U_X). \quad (6.81)$$

Elastic materials for which the stress can be derived from the Helmholtz free energy function are called *hyperelastic*.

The usual way the free energy function method is employed starts with using experimental observations to determine the functional form of the stress. With this one then shows there is a free energy function that will produce the given stress function. This is the approach used in the following examples.

*Example 1* For a linearly elastic material,  $T = EU_X$ . According to (6.81), the free energy function must satisfy  $R_0\Psi'(\epsilon) = E\epsilon$ . Integrating this expression we obtain  $R_0\Psi(\epsilon) = \frac{1}{2}E\epsilon^2$ . The constant of integration has not been included here as it has no impact on the stress function.  $\blacksquare$

*Example 2* If  $T = a(e^{b\epsilon} - 1)$ , where  $b \neq 0$ , then integrating  $R_0\Psi'(\epsilon) = a(e^{b\epsilon} - 1)$  yields  $R_0\Psi(\epsilon) = a\left(\frac{1}{b}e^{b\epsilon} - \epsilon\right)$ .  $\blacksquare$

*Example 3* For a viscous fluid it is assumed that  $\tau = \tau(\rho, v_x)$ . Assuming  $\psi$  depends on the same quantities, the Clausius-Duhem inequality (6.77) takes the form

$$\frac{\partial v}{\partial x} \left( \tau + \rho^2 \frac{\partial \psi}{\partial \rho} \right) - \rho \frac{\partial \psi}{\partial v_x} \frac{Dv_x}{Dt} \geq 0. \quad (6.82)$$

To obtain this result, the continuity equation  $\rho_t = -v\rho_x - \rho v_x$  has been used. The above inequality must hold if  $\tau = -\rho^2 \partial_\rho \psi$ . Since the other term does not depend on  $\tau$ , it follows that  $\partial_{v_x} \psi D_t v_x \leq 0$ . Since  $D_t v_x$  can be positive or negative, independently of the value of  $v_x$ , it must be that

$$\frac{\partial \psi}{\partial v_x} = 0. \quad (6.83)$$

Consequently, even though the stress might depend on  $v_x$ , the Clausius-Duhem inequality shows that the free energy function does not. Now, for a linear viscous model it is assumed that  $\tau + \rho^2 \psi_\rho = \alpha + \beta v_x$ , where  $\alpha$  and  $\beta$  are constants. Substituting this into (6.82), and making use of (6.83), we obtain  $(\alpha + \beta v_x)v_x \geq 0$ . This must hold for all values of  $v_x$ , and from this we conclude that  $\alpha = 0$  and  $\beta \geq 0$ . Setting  $p = \rho^2 \psi_\rho$ , then the resulting constitutive law for the stress is

$$\tau = -p + \beta \frac{\partial v}{\partial x}. \quad (6.84)$$

The function  $p$  is the pressure and the constant  $\beta$  is the viscosity. This means that for a viscous fluid there is an additional function to determine, and that is the pressure. This requires an additional equation, and for compressible fluids this is done by prescribing an equation of state. As an example, for an ideal gas it is assumed that  $p = a\rho^\gamma$ . ■

The thermodynamic foundations of continuum mechanics have been introduced only in the briefest terms, just enough to obtain the reduced entropy inequality (6.77). The fact is, this is a rich area, one that has generated more than its share of challenging mathematical and physical questions. For those who might want to learn more about this subject, the source for this material, and one that is oddly entertaining, is Truesdell (1984). In fact, the review of this book by Aris (1987) is also recommended.

As a final comment, if you are a bit uncertain about the meaning of entropy, you are in good company as even physicists are not in agreement (Swendsen, 2011). It is also interesting that even thermodynamic certainties, such as that absolute temperature cannot be negative, are now being questioned (Abraham and Penrose, 2017; Swendsen, 2018).

### 6.10.3 Hyperelasticity

As stated earlier, a hyperelastic material is one for which there is a Helmholtz free energy function  $\psi$  that is a function of the strain  $\epsilon = \frac{\partial U}{\partial X}$ . After working through the above example one might wonder if it is really necessary to introduce this idea. After all, the stress function can be deduced directly from the experimental data. As long as it is assumed that  $T$  depends on  $\epsilon$ , then the Principle of Dissipation is satisfied. The reason for this is that once  $T(\epsilon)$  is known you just integrate to find  $\psi$ , and this automatically guarantees the Principle of Dissipation is satisfied. Although this observation has merit, there are several reasons why the energy method is worth considering. One, very significant, reason is that in three-dimensional problems the integration method does not work except if the stress depends on the strain in a particular way.

Another reason for introducing the energy formulation relates to the mathematical problem derived from the constitutive law. If  $T$  is a nonlinear function of strain, then the momentum equation (6.50) is a nonlinear partial differential equation for the displacement. We saw in the last chapter how difficult it can be to determine whether a nonlinear equation has a solution, or whether it has just one solution. The energy formulation helps answer these questions, and this is illustrated in the next example.

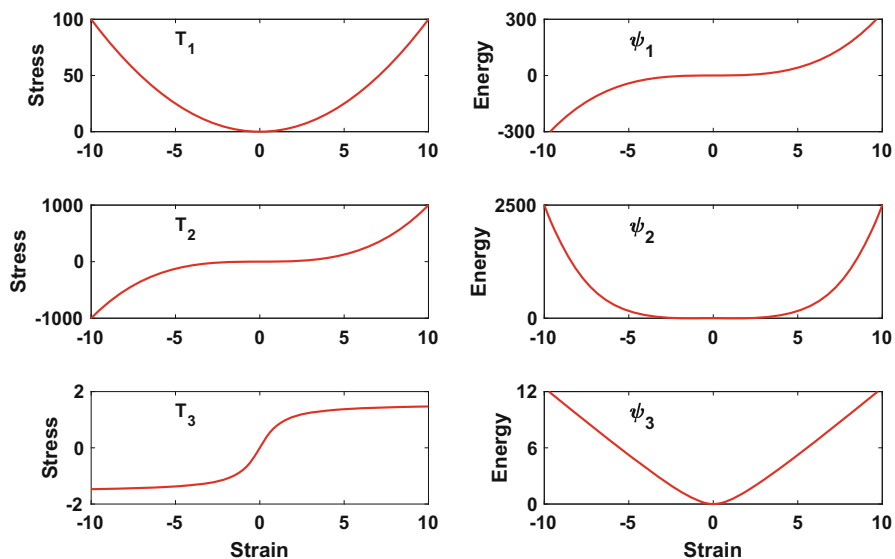
*Example (Bungie Cord Revisited)* For the bungie cord example we solved the momentum equation to find the stress, given in (6.53). To determine the displacement of the cord the linear elastic constitutive law was used. Suppose, instead, the material is nonlinear. We will consider three different nonlinear stress-strain laws, along with their corresponding free energy functions:

$$T_1 = E\epsilon^2, \quad \psi_1 = \frac{1}{3}E\epsilon^3, \quad (6.85)$$

$$T_2 = E\epsilon^3, \quad \psi_2 = \frac{1}{4}E\epsilon^4, \quad (6.86)$$

$$T_3 = E \arctan(\epsilon), \quad \psi_3 = E \left( \epsilon \arctan(\epsilon) - \frac{1}{2} \ln(1 + \epsilon^2) \right). \quad (6.87)$$

These functions are plotted in Fig. 6.12 in the case of when  $E = 1$ . The strain interval in this figure is larger than what is possible physically, but is used to help make the points to follow. For the problem at hand the question is, given the stress,



**Fig. 6.12** Nonlinear stress-strain functions and their corresponding Helmholtz free energy functions

can we uniquely determine the strain? For each stress function we have the following observations:

- $T_1$ : Given a value for the stress, other than zero, there are two possible values of the strain if  $T_1 > 0$ , and no strain values when  $T_1 < 0$ . In other words, except for zero, there is no solution or else the solution is not unique.
- $T_2$ : Given any value for the stress, there is a unique strain. In other words, there is a solution and it is unique. Note that the free energy for this stress function is concave up, or equivalently, convex.
- $T_3$ : For each stress value there is a unique strain. However, there are stress values, such as  $T_3 = 2$ , for which there is no corresponding strain. In other words, if there is a solution it is unique, but there are stress values for which there is no solution. Note that the free energy for this stress function is convex. ■

In general, to prevent multiple strain values as happened with  $T_1$ , but not with  $T_2$  and  $T_3$ , the stress-strain law must be strictly monotonic increasing. This translates into the requirement that  $\psi$  is a strictly convex function of the strain, and this occurs if

$$\frac{d^2\psi}{d\epsilon^2} > 0. \quad (6.88)$$

However, this assumption is not enough. As shown with  $T_3$ , to guarantee that a solution exists, the stress values must have the right limiting behavior. This is the extreme value issue that was discussed earlier. Given that the strain interval is  $-1 < \epsilon < \infty$ , the specific requirement for this one-dimensional problem is that

$$\lim_{\epsilon \rightarrow -1} T = -\infty \quad \text{and} \quad \lim_{\epsilon \rightarrow \infty} T = \infty. \quad (6.89)$$

This means that it is assumed that it takes infinite energy to expand a finite bar to one of infinite length, and it also takes infinite energy to compress a bar down to one with zero length. None of the above three energy functions satisfies the  $\epsilon \rightarrow -1$  condition, but examples of those that do can be found in Exercises 6.25 and 6.29.

*Example 1* For a linearly elastic material,  $R_0\Psi(\epsilon) = \frac{1}{2}E\epsilon^2$ . From the convexity condition, it is required that  $E \geq 0$ . ■

*Example 2* If  $T = a(e^{b\epsilon} - 1)$ , then  $R_0\Psi(\epsilon) = a\left(\frac{1}{b}e^{b\epsilon} - \epsilon\right)$ . From the convexity condition, it is required that  $ab \geq 0$ . ■

The requirements for multidimensional problems are harder to determine. For example, when there is more than one spatial dimension, it has been shown that a free energy function that is convex will not be frame-indifferent. This was the motivation for introducing a milder form of convexity, something called polyconvexity. This is beyond the scope of this textbook, and the interested reader should consult Marsden and Hughes (1994) for further details.

## Exercises

### Section 6.2

**6.1** Assume the motion is described by  $\mathcal{X}(X, t) = Xe^t$ .

- Consider the cross-section that at  $t = 5$  is located at  $x = 1$ . Where was it at  $t = 0$ ?
- Find  $u(x, t)$  and  $U(X, t)$ .
- Find  $v(x, t)$  and  $V(X, t)$ .
- What is the velocity of the cross-section that is at  $x = 2$  at time  $t$ ? What is the velocity at time  $t$  of the cross-section that starts at  $x = 2$ ?
- Suppose the temperature of the bar is  $\theta(x, t) = x^5 + 4t$ . What is the rate of change of  $\theta$  following a material section?

**6.2** Suppose that at  $t = 0$  the bar occupies the interval  $0 \leq X \leq 1$  and the motion of the bar is governed by the equation  $\mathcal{X}(X, t) = X + Xt^2$ .

- What spatial interval does the bar occupy at  $t = 2$ ?
- Find  $V(X, t)$ . What are the limits on  $X$ ?
- Find  $v(x, t)$ . What are the limits on  $x$ ?
- Suppose the temperature of the bar is  $\theta(x, t) = xt^3$ . What is the rate of change of  $\theta$  following a material section?

**6.3** This problem considers how the displacement can be determined from the velocity when using spatial coordinates. Therefore, in this problem, it is assumed that  $v(x, t)$  is known.

- The direct approach to finding  $u$  uses (6.14). Show that this leads to a first-order partial differential equation for  $u$ . What is the initial condition for  $u$ ?
- Another approach involves first converting to material coordinates. Explain why this results in having to solve

$$\frac{\partial U}{\partial t} = v(U + X, t),$$

where  $U(X, 0) = 0$ . Once  $U(X, t)$  is known, explain how to determine  $u(x, t)$ .

- Using the approach in either part (a) or part (b), find  $u$  if  $v = x/(\alpha + t)$ , where  $\alpha$  is a positive constant.

**6.4** This problem explores the transformation between the material and spatial coordinate systems.

- Use the impenetrability of matter assumption to prove that  $\partial_X U \geq -1$ .
- Give an example of an  $U$  so that (6.18) is satisfied but there is at least one point where  $\partial_X U = -1$ .



- (c) Explain why the points where  $\partial_X U = -1$  must be isolated. Specifically, show that if there is an interval  $X_L < X < X_R$  where  $\partial_X U = -1$ , then (6.18) does not hold.
- (d) Show that if (6.19) holds, then so does (6.20).

**6.5** In this problem assume, as usual, that  $f(x, t)$  and  $g(x, t)$  are smooth functions.

- (a) Show that

$$(1) \quad \frac{D(f+g)}{Dt} = \frac{Df}{Dt} + \frac{Dg}{Dt}, \quad \text{and} \quad (2) \quad \frac{D(fg)}{Dt} = f \frac{Dg}{Dt} + g \frac{Df}{Dt}.$$

- (b) Explain why it is not necessarily true that  $\frac{D}{Dt} \frac{\partial f}{\partial t} = \frac{\partial}{\partial x} \frac{Df}{Dt}$ . What about the equation  $\frac{D}{Dt} \frac{\partial f}{\partial t} = \frac{\partial}{\partial t} \frac{Df}{Dt}$ ?
- (c) If  $h = h(x)$ , then show that  $\frac{D}{Dt} h(f) = h'(f) \frac{Df}{Dt}$ .

**6.6** Derive the following.

- |                                                                                                                          |                                                                        |
|--------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|
| (a) $v = \frac{u_t}{1 - u_x}$ .                                                                                          | (d) $\frac{\partial V}{\partial X} = \frac{v_x}{1 - v_x}$ .            |
| (b) $\frac{\partial u}{\partial x} = \frac{U_x}{1 + U_X}$ .                                                              | (e) $\frac{\partial^2 U}{\partial X^2} = \frac{u_{xx}}{(1 - u_x)^3}$ . |
| (c) $\frac{\partial v}{\partial x} = \frac{\partial}{\partial t} \ln \left( 1 + \frac{\partial U}{\partial X} \right)$ . | (f) $\frac{D}{Dt} u_x = (1 - u_x) v_x$ .                               |

**6.7** The deformation gradient, in material coordinates, is defined as  $F(X, t) = \frac{\partial \mathcal{X}}{\partial X}$ . This function is used extensively in continuum mechanics when studying nonlinear elastic materials.

- (a) Show that  $F(X, t) = 1 + \frac{\partial U}{\partial X}$ , and  $F(X, 0) = 1$ .
- (b) Letting  $f(x, t)$  denote the deformation gradient in spatial coordinates show that  $\frac{Df}{Dt} = \frac{\partial v}{\partial x} f$ .
- (c) The function  $C(X, t) = F^2$  is known as the Cauchy-Green deformation tensor in material coordinates. Letting  $c(x, t)$  denote this function in spatial coordinates show that  $\frac{Dc}{Dt} = 2 \frac{\partial v}{\partial x} f^2$ .

**6.8** This problem considers some of the restrictions on the displacement function.

- (a) Why is it not possible to have  $\mathcal{X}(X, t) = \frac{1}{2} X \cos(t)$ ?
- (b) Why is it not possible to have, at any given value of  $t$ ,  $u(0, t) = -1$  and  $u(1, t) = 1$ ?
- (c) Why is it not possible to have  $U(0, t) = 1$  and  $U(1/2, t) = 0$ ?
- (d) Prove that if  $X_1 < X_2$ , then  $U(X_1, t) < X_2 - X_1 + U(X_2, t)$ .
- (e) Explain why it is not possible to have a displacement function of the form  $U = \alpha \sin(X)$ , where  $\alpha > 1$ .

## Sections 6.3–6.6

**6.9** In the derivation of the equations of motion it was assumed that the cross-sectional area is constant. This problem examines what happens if this assumption is dropped and  $\sigma = \sigma(x)$ .

- (a) Derive the resulting continuity equation in spatial coordinates, and then show that the material coordinates version is

$$R(X, t) = \frac{R_0}{1 + U_X} \frac{\sigma(X)}{\sigma(X + U)}.$$

- (b) Derive the momentum equation in spatial coordinates, and then show that the material coordinates version is

$$R_0 \sigma(X) \frac{\partial^2 U}{\partial t^2} = R_0 \sigma(X) F + \frac{\partial}{\partial X} (\sigma(X) T).$$

Note that in the derivation you will need to use that  $\sigma(x)\tau = \sigma(X)T$ . This equality comes from the assumption that the value of the force between cross-sections is the same in the material and spatial systems. Consequently, since  $\tau$  uses the current area  $\sigma(x)$  and  $T$  use the original area  $\sigma(X)$  it follows that  $\sigma(x)\tau = \sigma(X)T$ .

- (c) Assuming  $F = 0$ , show that the steady-state solution of the momentum equation is

$$T = \frac{f_0}{\sigma(X)},$$

where  $f_0$  is a constant.

- (d) Assuming material linearity, show that the momentum equation in part (b) can be written as

$$\frac{\partial^2 U}{\partial t^2} = c^2 \frac{\partial^2 U}{\partial X^2} + c^2 \frac{1}{\sigma} \frac{\partial \sigma}{\partial X} \frac{\partial U}{\partial X} + F,$$

where  $c$  is a positive constant. This is known as Webster's equation, or Webster's horn equation. It gets this name because it arises in the study of acoustic waves in a horn or loudspeaker.

**6.10** Instead of using a material volume to derive the equations of motion, it is possible to use a fixed spatial region. This is the control volume approach used to derive the traffic flow equation in the previous chapter.

- (a) Given a spatial location  $x_0$ , consider the interval  $x_0 - \Delta x \leq x \leq x_0 + \Delta x$ . Explain where each term in the following equation comes from:

$$\begin{aligned} \int_{x_0 - \Delta x}^{x_0 + \Delta x} \sigma \rho(x, t) a(x, t) dx \\ = \tau(x_0 + \Delta x, t) \sigma - \tau(x_0 - \Delta x, t) \sigma + \int_{x_0 - \Delta x}^{x_0 + \Delta x} \sigma \rho(x, t) f(x, t) dx, \end{aligned}$$

where  $a(x, t)$  is the acceleration.

- (b) Assuming small  $\Delta x$ , show that the equation in part (a) reduces to

$$2\sigma \Delta x \rho(x_0, t) a(x_0, t) = 2\sigma \Delta x \tau_x(x_0, t) + 2\sigma \Delta x \rho(x_0, t) f(x_0, t) + O(\Delta x^2).$$

- (c) Using the result from part (a), derive the momentum equation.

**6.11** This problem derives general forms of the balance law, using the same notations used in (6.22), (6.31), and (6.73). With this in mind, let  $f(x, t)$  be a quantity that is measured per unit volume.

- (a) Explain where each term in the following balance equation comes from:

$$\frac{d}{dt} \int_{\alpha(t)}^{\beta(t)} \sigma f dx = -\sigma J \Big|_{x=\alpha}^{x=\beta} + \int_{\alpha(t)}^{\beta(t)} \sigma Q dx.$$

- (b) Identify the functions  $f$ ,  $J$ , and  $Q$  for the equations (6.22), (6.31), and (6.73).  
 (c) Show that the balance law in part (a) reduces to

$$\frac{\partial f}{\partial t} + \frac{\partial(vf)}{\partial x} = -\frac{\partial J}{\partial x} + Q.$$

**6.12** Transform the initial conditions  $U|_{t=0} = G(X)$  and  $V|_{t=0} = H(X)$  into initial conditions for  $u$  and  $u_t$ .

## Sections 6.7 and 6.8

**6.13** In three dimensions it is more common to use the Green strain, and this problem explores some of the differences between it and the Lagrangian strain.

- (a) Rewrite the Lagrangian and Green ratios listed in Table 6.4 in terms of  $\lambda = \ell/\ell_0$  and then on the same axes, sketch each ratio for  $0 < \lambda < \infty$ .  
 (b) Derive the formula for the Green strain  $\epsilon_g$ .  
 (c) In the case of when  $U_X$  is close to zero, explain why the Green strain reduces to the Lagrangian strain.

- (d) Under what circumstances would it be more appropriate to assume  $T = E\epsilon_g$  rather than  $T = E\epsilon$ ?
- (e) What is the resulting equation of motion for the displacement  $U$  if one assumes  $T = E\epsilon_g$ ?

**6.14** This exercise examines the Hencky strain given in Table 6.4.

- (a) Rewrite the Hencky ratio in terms of  $\lambda = \ell/\ell_0$  and then sketch the ratio for  $0 < \lambda < \infty$ .
- (b) Derive the formula for the Hencky strain from the ratio.
- (c) What is the resulting equation of motion for the displacement  $U$  if one assumes  $T = E\epsilon_h$ ?

**6.15** This exercise examines the midpoint strain given in Table 6.4.

- (a) Rewrite the midpoint ratio in terms of  $\lambda = \ell/\ell_0$  and then sketch the ratio for  $0 < \lambda < \infty$ .
- (b) Derive the formula for the midpoint strain from the ratio.
- (c) What is the resulting equation of motion for the displacement  $U$  if one assumes  $T = E\epsilon_m$ ?

**6.16** In modeling rubber as a chained polymer using what is known as a fixed junction model it is determined that two useful strain measures are  $\lambda^2 - 1/\lambda$  and  $\lambda - 1/\lambda^2$ , where  $\lambda = \ell/\ell_0$  in the extension ratio.

- (a) On the same axes, sketch each strain measure for  $0 < \lambda < \infty$ .
- (b) Derive the strain for each strain measure.
- (c) Does either strain in part (b) reduce to  $U_X$  if  $U_X$  is small?

**6.17** Suppose in the bungee cord example the initial density is not constant, and  $R(X, 0) = \alpha(1 + X/L_0)$ . What is the steady-state length  $L$  of the bungee cord?

**6.18** If someone is attached to the end of the bungee cord, the boundary condition at  $X = L_0$  changes to  $T = T_0$ , where  $T_0$  is a given positive constant. What is the steady-state length  $L$  of the bungee cord?

**6.19** For the space elevator, a cable extends radially into space, with one end attached to the Earth and the other end, out in space, is stress free. Assume that initially the cable has length  $L_0$  and a constant density  $R_0$ .

- (a) The body forces include gravity  $F_g = -g(R/r)^2$  and a centripetal force  $F_c = \omega^2 r$ , where  $R$  and  $\omega$  are the equatorial radius and angular rotational velocity, respectively, of the Earth. The resulting body force is  $F = F_g + F_c$ . In substituting this into the momentum equation (6.42) explain why it is necessary to take  $r = R + X + U$ .
- (b) Assuming a steady state, so  $T$  and  $U$  are independent of  $t$ , write down the resulting momentum equation. What is the boundary condition at  $X = 0$ ? What is the boundary condition at  $X = L_0$ ?
- (c) Assuming  $U$  is small relative to  $R$ , a first order approximation to the problem you derived in part (b) can be obtained by replacing the  $R + X + U$  terms

with  $R + X$ . Write down the resulting momentum equation and the boundary condition at  $X = L_0$ .

- (d) Solve the problem in part (c) for  $T$ .
- (e) The values for the parameters appearing in this problem are (approximately):  $\omega = 7.3 \times 10^{-5}/\text{s}$ ,  $R = 6400 \text{ km}$ ,  $g = 9.8 \text{ m/s}^2$ ,  $L_0 = 10^5 \text{ km}$ ,  $R_0 = 1300 \text{ kg/m}^3$ , and  $E = 60 \text{ GPa}$  (the latter two values are representative of multi-walled carbon nanotubes). What is the stress the cable puts on the base where it is attached to the Earth?
- (f) Assuming the cable is linearly elastic, use your result in part (d) to find the length of the cable at steady state.
- (g) Using the values given in part (e), determine the steady-state length  $L$  of the cable.

**6.20** A linearly elastic bar is made of two different materials and before being stretched it occupies the interval  $0 \leq X \leq L_0$ . Also, before being stretched, for  $0 \leq X < X_0$  the modulus and density are  $E = E_L$  and  $R = R_L$ , while for  $X_0 < X \leq L_0$  they are  $E = E_R$  and  $R = R_R$ . Both  $R_L$  and  $R_R$  are constants.

- (a) The requirements at the interface, where  $X = X_0$ , are that the displacement and stress are continuous. Express these requirements mathematically using one-sided limits.
- (b) Suppose the bar is stretched and the boundary conditions are  $U(0, t) = 0$  and  $U(L_0, t) = L - L_0$ . Assume there are no body forces. Find the steady-state solution for the density, displacement and stress.

**6.21** This problem concerns the constitutive law  $T(\epsilon) = E \ln(1 + \epsilon)$ , where  $\epsilon$  is given in (6.49).

- (a) Sketch  $T$  for  $-1 < \epsilon < \infty$ . Also sketch, on the same axis, the function  $E\epsilon$ .
- (b) Will the bungee cord stretch farther if  $T(\epsilon) = E \ln(1 + \epsilon)$  than it does when  $T(\epsilon) = E\epsilon$ ? Use your sketch from part (a) to answer this question.
- (c) For the bungee cord example, and using  $T(\epsilon) = E \ln(1 + \epsilon)$ , find  $L$ .

## Section 6.9

**6.22** This problem explores some additional ideas related to the morphological basis for deformation of a metal.

- (a) The binding energy  $V_0 = V(r_0)$ , where  $V$  is given in (6.63) and  $r_0$  in (6.64), is the minimum energy needed to break the atomic bonds. Show that

$$V = \frac{V_0}{m-n} \left[ -n \left( \frac{r_0}{r} \right)^m + m \left( \frac{r_0}{r} \right)^n \right].$$

(b) Show that

$$E \approx -\frac{mnV_0}{r_0^3}.$$

One conclusion that comes from this result is that materials with a high binding energy, and a small interatomic spacing, have a relatively large Young's modulus.

**6.23** This problem examines the elastic modulus when the interatomic forces are described using the Morse potential function, which is

$$V = \beta \left( e^{-2\alpha(r-r_0)} - 2e^{-\alpha(r-r_0)} \right),$$

where  $\alpha$  and  $\beta$  are positive constants.

- Show that  $V'(r_0) = 0$ .
- What is the resulting force function  $F$ ? Identify the term accounting for the repulsive component of the force, and the term responsible for the attractive component.
- Sketch  $V$  and  $F$  for  $0 < r < \infty$ . Comment on the qualitative differences of these functions compared to those shown in Fig. 6.9.
- What is the resulting approximation for the elastic modulus?
- It has been found that for carbon nanotubes,  $\beta = 3.77 \text{ eV}$ ,  $\alpha = 26.25 \text{ nm}^{-1}$ , and  $r_0 = 0.14 \text{ nm}$  (Liew et al. 2005). Use these values to estimate the Young's modulus. Note that  $1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}$ .

## Section 6.10

**6.24** This problem considers if various constitutive laws satisfy the Principle of Material Frame-Indifference.

- Show that  $\frac{\partial v}{\partial t}$  is not frame-indifferent. Explain why this shows that  $\tau = \tau(v_t)$  does not satisfy the Principle of Material Frame-Indifference.
- Does  $T = T(V_t)$  satisfy the Principle of Material Frame-Indifference?
- Show that  $\tau = \tau(u_x, u_{xt})$  satisfies the Principle of Material Frame-Indifference.
- Does  $T = T(U_X, U_{Xt})$  satisfy the Principle of Material Frame-Indifference?

**6.25** The Mooney-Rivlin model for rubber assumes  $T = (\alpha + \frac{\beta}{\lambda})(\lambda^2 - \frac{1}{\lambda})$ , where  $\lambda = 1 + \epsilon$ , and  $\alpha, \beta$  are positive constants.

- Sketch the stress for  $-1 < \epsilon < \infty$ .
- By assuming  $\epsilon$  is close to zero, determine how  $\alpha$  and  $\beta$  are related to Young's modulus.
- Find the Helmholtz free energy function  $\Psi$ .

**6.26** Suppose it is assumed that  $\Psi = \Psi(\epsilon_g)$ , where  $\epsilon_g = U_X + \frac{1}{2}U_X^2$  is the Green strain.

- (a) Given that  $U_X > -1$ , sketch  $\epsilon_g$  as a function of  $U_X$ .
- (b) Show that  $T = R_0(1 + U_X)\Psi'(\epsilon_g)$ .
- (c) Suppose it is known that  $T = E\epsilon_g$ . Use this to show that the free energy function is

$$\Psi(\epsilon_g) = \frac{E}{3R_0}(\epsilon_g - 1)\sqrt{1 + 2\epsilon_g}.$$

**6.27** Suppose it is assumed that  $\psi = \psi(\epsilon_a)$ , where  $\epsilon_a = u_x - \frac{1}{2}u_x^2$  is the Almansi strain.

- (a) Given that  $u_x < 1$ , sketch  $\epsilon_a$  as a function of  $u_x$ .
- (b) Show that  $\frac{D}{Dt}\psi = (1 - u_x)^2 v_x \psi'(\epsilon_a)$ .
- (c) Show that  $\tau = \rho(1 - u_x)^2 \psi'(\epsilon_a)$ .
- (d) Suppose it is known that  $\tau = E\epsilon_a$ . What is the free energy function?

**6.28** The mechanical energy equation is

$$\rho \frac{D}{Dt} \left( \frac{1}{2} v^2 \right) + \tau \frac{\partial v}{\partial x} = \frac{\partial}{\partial x} (v\tau) + \rho v f.$$

- (a) Derive this directly from the momentum equation.
- (b) Combine the result from part (a) with (6.74) to show that

$$\rho \frac{D}{Dt} \left( \frac{1}{2} v^2 + \chi \right) = \frac{\partial}{\partial x} (v\tau) + \rho v f.$$

The above equation represents the time rate of change of the total energy balanced with the energy flux associated with the stress and the rate of work of the body forces.

**6.29** An energy function used for biological tissue is

$$\Psi = \frac{\alpha}{\lambda^{2\beta}} e^{\beta\lambda^2},$$

where  $\lambda = 1 + \epsilon$ , and  $\alpha, \beta$  are positive constants (Holmes, 1986).

- (a) Show that

$$T = \frac{1}{2} E \frac{\lambda^2 - 1}{\lambda^{2\beta+1}} e^{\beta(\lambda^2-1)},$$

where  $E$  is a positive constant.

- (b) Show that if  $\epsilon$  is small, then the formula in part (a) reduces to the linearly elastic constitutive law given in (6.51).
- (c) Show that  $T$  is a strictly monotonic increasing function of  $\lambda$ . Explain why this means that  $T$  is a strictly monotonic increasing function of  $\epsilon$ .
- (d) Show that  $T$  satisfies the limit conditions in (6.89).



# Chapter 7

## Elastic and Viscoelastic Materials



### 7.1 Linear Elasticity

A particularly successful application of continuum mechanics is linear elasticity. For a linearly elastic material, the constitutive law for the stress is

$$T = E \frac{\partial U}{\partial X}, \quad (7.1)$$

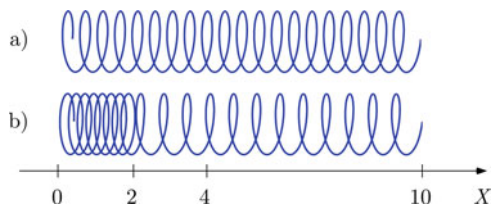
where  $E$  is Young's modulus. The momentum equation (6.42) in this case reduces to

$$\frac{\partial^2 U}{\partial t^2} = c^2 \frac{\partial^2 U}{\partial X^2} + F, \quad (7.2)$$

where  $c^2 = E/R_0$ . It is assumed that both  $E$  and  $R_0$  are constants. Therefore, the equation of motion for a linearly elastic material is a wave equation for the displacement. One of the objectives of this chapter is to solve this equation, and then use the solution to understand how an elastic material responds.

It is important to point out that the linear elastic model we are considering comes from assuming that the stress is a linear function of the Lagrangian strain (6.49). As is evident from Fig. 6.5, exactly what strains this is valid for depends on the specific material under study. Also, if one of the other strains listed in Table 6.4 is used, a linear constitutive law for the stress does not lead to a linear momentum equation as happens in (7.2). This observation will be reconsidered later when discussing what is known as the assumption of geometric linearity.

There is a long list of methods that can be used to solve problems in linear elasticity, and this includes separation of variables, Green's functions, Fourier



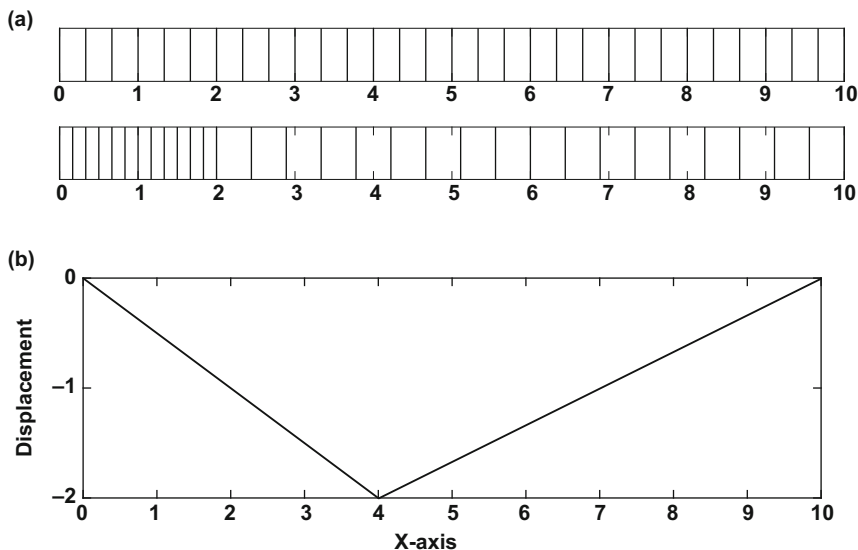
**Fig. 7.1** (a) A slightly extended slinky (spring) is held at  $X = 0$  and at  $X = 10$ . (b) The loop that was at  $X = 4$  is moved over to  $X = 2$ , producing a compression in the region  $0 \leq X < 2$ , and an expansion in  $2 < X \leq 10$

transforms, Laplace transforms, and the method of characteristics. The latter two will be used in this chapter, and the reasons for this will be explained as the methods are developed. Before doing this we consider a more basic issue, and this has to do with the form of the mathematical solution and its connection to the physical problem.

*Example (Stretching a Rubber Band)* Suppose a rubber band is stretched a small amount with one end held at  $X = 0$  and the other end held at  $X = 10$ . One then moves the cross-section at  $X = 4$  to  $X = 2$ . This situation is illustrated in Fig. 7.1 for a slinky, which is not exactly a rubber band but behaves in a similar manner. For the spring, the distance between the loops is a measure of the strain. As an example, in Fig. 7.1b, the loops in  $0 \leq X < 2$  and in  $4 < X \leq 10$  are both uniformly spaced, indicating a uniform strain in these two regions. The fact that the loops in  $0 \leq X < 2$  are closer together than they are in the upper figure indicates a constant compressive strain. For a similar reason there is a constant tensile strain in  $4 < X \leq 10$ . Returning to the rubber band, we will assume that at rest it can be modeled as a linearly elastic material. To satisfy the given boundary conditions, it is required that the displacement satisfy  $U = 0$  at  $X = 0$  and at  $X = 10$ . Also, given that the cross-section that was at  $X = 4$  is moved over to  $X = 2$ , then it is required that  $U = -2$  at  $X = 4$ . From (7.2), the steady state with  $F = 0$  means that  $U_{XX} = 0$ . So,  $U$  is a linear function of  $X$ . More precisely, it is linear for  $0 < X < 4$ , and it is another linear function for  $4 < X < 10$ . For  $0 < X < 4$ , the linear function that satisfies  $U(0) = 0$  and  $U(4) = -2$  is  $U = -X/2$ . For  $4 < X < 10$ , the linear function that satisfies  $U(10) = 0$  and  $U(4) = -2$  is  $U = (X - 10)/3$ . We therefore have the piecewise linear solution

$$U = \begin{cases} -X/2 & \text{if } 0 \leq X \leq 4, \\ (X - 10)/3 & \text{if } 4 \leq X \leq 10. \end{cases} \quad (7.3)$$

The conventional method for plotting such a function is given in Fig. 7.2b. It shows, for example, that the point that started at  $X = 4$  moves in the negative direction to  $X = 2$ . Although there is nothing wrong with this plot, it obfuscates what is happening in the rubber band and seems to have no connection with what is illustrated in Fig. 7.1. Another method for plotting the solution is given in Fig. 7.2a.



**Fig. 7.2** The rubber band at rest example. In (a) the upper bar shows evenly spaced cross-sections in the rubber band before it is pulled, and the lower bar shows where they are located after it is pulled. In (b) the displacement (7.3) is plotted in a more traditional method

The upper bar shows cross-sections equally spaced along the rubber band, before the rubber band is pulled. In the lower bar in Fig. 7.2a the positions of the same cross-sections are shown after the rubber band has been pulled. The position of any given cross-section is  $\mathcal{X} = X + U$ , where  $U$  is given in (7.3). What is seen is that the cross-sections that started out uniformly spaced in  $0 \leq X \leq 4$  end up uniformly spaced in the interval  $0 \leq X \leq 2$ . The difference is that they are closer together due to the fact that the rubber band is being compressed in this region. In contrast, the cross-sections that started out in  $4 \leq X \leq 10$  get farther apart after pulling, and this is due to the stretching of the rubber band in this region. ■

The solution in the rubber band example illustrates some general characteristics that arise in elasticity. Whenever the strain is negative, so  $U_X < 0$ , the cross-section is said to be in compression. This means that the cross-sections in this vicinity are closer together than they were before the load was applied. In contrast, whenever the strain is positive the cross-section is in tension. In Fig. 7.2a, the cross-sections that start out in  $4 < X < 10$  end up in  $2 < X < 10$ , and are therefore in tension because  $U_X = 1/3$ . Similarly, those that start out in  $0 < X < 4$  end up in  $0 < X < 2$ , and they are in compression because  $U_X = -1/2$ .

### 7.1.1 Method of Characteristics

Suppose the bar is very long, so it is reasonable to assume  $-\infty < X < \infty$ . Also, it is assumed that there are no body forces. The initial conditions that will be used are

$$U(X, 0) = f(X), \quad \partial_t U(X, 0) = g(X). \quad (7.4)$$

With the given assumptions, the wave equation (7.2) can be written as

$$\left( \frac{\partial^2}{\partial t^2} - c^2 \frac{\partial^2}{\partial X^2} \right) U = 0.$$

Factoring the derivatives, the equation takes the form

$$\left( \frac{\partial}{\partial t} - c \frac{\partial}{\partial X} \right) \left( \frac{\partial}{\partial t} + c \frac{\partial}{\partial X} \right) U = 0. \quad (7.5)$$

Our goal is to change coordinates, from  $(X, t)$  to  $(r, s)$ , so the above equation can be written as

$$\frac{\partial}{\partial r} \left( \frac{\partial U}{\partial s} \right) = 0. \quad (7.6)$$

What we want, therefore, is the following:

$$\frac{\partial}{\partial r} = \frac{\partial}{\partial t} - c \frac{\partial}{\partial X}, \quad (7.7)$$

$$\frac{\partial}{\partial s} = \frac{\partial}{\partial t} + c \frac{\partial}{\partial X}. \quad (7.8)$$

To determine how this can be done assume  $X = X(r, s)$ ,  $t = t(r, s)$ . In this case, using the chain rule

$$\frac{\partial}{\partial r} = \frac{\partial X}{\partial r} \frac{\partial}{\partial X} + \frac{\partial t}{\partial r} \frac{\partial}{\partial t}, \quad (7.9)$$

$$\frac{\partial}{\partial s} = \frac{\partial X}{\partial s} \frac{\partial}{\partial X} + \frac{\partial t}{\partial s} \frac{\partial}{\partial t}. \quad (7.10)$$

Comparing (7.7) and (7.9), we require  $\frac{\partial X}{\partial r} = -c$  and  $\frac{\partial t}{\partial r} = 1$ . Similarly, comparing (7.8) and (7.10), we require  $\frac{\partial X}{\partial s} = c$  and  $\frac{\partial t}{\partial s} = 1$ . Solving these equations gives us that  $X = c(-r + s)$  and  $t = r + s$ . Inverting this transformation one finds,

$$r = -\frac{1}{2c}(X - ct), \quad s = \frac{1}{2c}(X + ct). \quad (7.11)$$

This change of variables reduces the wave equation to (7.6). The general solution of this is  $U = F(r) + G(s)$  where  $F$  and  $G$  are arbitrary functions. Reverting back to  $X, t$ , and absorbing the  $\frac{1}{2c}$  into the arbitrary functions, we obtain the solution

$$U(X, t) = F(X - ct) + G(X + ct), \quad (7.12)$$

where  $F$  and  $G$  are determined from the initial conditions. With this we have that the general solution of the problem consists of the sum of two traveling waves. One, with profile  $F$ , moves to the right with speed  $c$ , and the other, with profile  $G$ , moves to the left with speed  $c$ .

It remains to have (7.12) satisfy the initial conditions (7.4). Working out the details, one finds that the solution is

$$U(X, t) = \frac{1}{2}f(X - ct) + \frac{1}{2}f(X + ct) + \frac{1}{2c} \int_{X-ct}^{X+ct} g(z)dz. \quad (7.13)$$

This is known as the *d'Alembert solution* of the wave equation. It is crystal clear from this expression how the initial conditions contribute to the solution. Specifically, the initial displacement  $f(X)$  is responsible for two traveling waves, both moving with speed  $c$  and traveling in opposite directions. The initial velocity  $g(X)$  contributes over an ever-expanding interval, the endpoints of this interval moving with speed  $c$ .

*Example* Suppose the initial conditions are  $U(X, 0) = f(X)$  and  $\partial_t U(X, 0) = 0$ , where  $f(X)$  is the rectangular bump

$$f(X) = \begin{cases} 1 & \text{if } -1 \leq X \leq 1, \\ 0 & \text{otherwise.} \end{cases} \quad (7.14)$$

From (7.13) the solution is

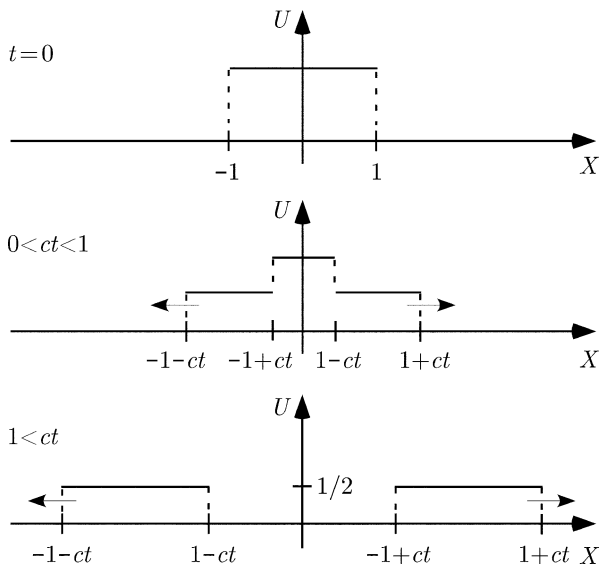
$$U(X, t) = \frac{1}{2}f(X - ct) + \frac{1}{2}f(X + ct). \quad (7.15)$$

This is shown in Fig. 7.3 and it is seen that the solution consists of two rectangular bumps, half the height of the original, traveling to the left and right with speed  $c$ .

■

The nice thing about the method of characteristics is that it produces a solution showing the wave-like nature of the response. Its flaw is that the derivation assumes that the interval is infinitely long. It is possible in some cases to use it on finite intervals, by accounting for the reflections of the waves at the boundaries. The mathematical representation of such a solution is obtained in the slinky example in the next section. For finite intervals other methods can be used. One is separation of variables, which is a subject often covered in elementary partial differential equation textbooks. Another is the Laplace transform, and this is the one pursued here.

**Fig. 7.3** Solution of the wave equation obtained using the d'Alembert solution (7.15)



### 7.1.2 Laplace Transform

Earlier, in Chap. 4, we used the Fourier transform to solve the diffusion equation. This could also be used on the wave equation, but the Laplace transform is used instead. One reason is that it is an opportunity to learn something new. Another reason is that the Laplace transform is particularly useful for cracking open some of the problems that will arise later in the chapter when studying viscoelasticity.

The Laplace transform of a function  $U(t)$  is defined as

$$\widehat{U}(s) \equiv \int_0^{\infty} U(t) e^{-st} dt. \quad (7.16)$$

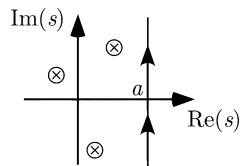
The conditions, and requirements, on  $U(t)$  and the Laplace variable  $s$  will be discussed in the following pages.

We will need to be able to determine  $U(t)$  given  $\widehat{U}(s)$ , and for this we need the inverse transform. It can be shown that if  $U$  is continuous at  $t$ , then

$$U(t) = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} \widehat{U}(s) e^{st} ds. \quad (7.17)$$

The integral here is a line integral in the complex plane, along the vertical line  $\text{Re}(s) = a$  (see Fig. 7.4). It is evident from the above line integral that the variable  $s$  in (7.16) is complex valued. A second observation is that the inverse transform (7.17) is not as simple as might be expected from (7.16). Although some of the more entertaining mathematical problems arise when inverting the Laplace transform

**Fig. 7.4** Contour used in the formula for the inverse Laplace transform (7.17). It must be to the right of any singularity of  $\widehat{U}$ , which are indicated using the symbol  $\otimes$



using contour integration in the complex plane, most people rely on tables. This will be the approach used here, and we will mostly determine the inverse using the relatively small collection of formulas listed in Table 7.1.

It is convenient to express the Laplace transform in operator form, and write (7.16) as  $\widehat{U} = \mathcal{L}(U)$ . Using this notation, the inverse transform (7.17) is  $U = \mathcal{L}^{-1}(\widehat{U})$ . It should be restated that the inverse formula assumes that  $U$  is continuous at  $t$ . If it is not, and  $U$  has a jump discontinuity at  $t$ , then the right-hand side of (7.17) equals the average of the jump. This means that, for  $t > 0$ ,

$$\mathcal{L}^{-1}(\widehat{U}) = \frac{1}{2} (U(t^+) + U(t^-)). \quad (7.18)$$

This result, that one obtains the average of the function at a jump, is consistent with what was found for the inverse Fourier transform in Chap. 4.

A particularly important property of the Laplace transform, and its inverse, is linearity. Specifically, given functions  $U(t)$  and  $V(t)$ , along with constants  $a$  and  $b$ , then

$$\mathcal{L}(aU + bV) = a\mathcal{L}(U) + b\mathcal{L}(V).$$

Similarly, given transforms  $\widehat{U}(s)$  and  $\widehat{V}(s)$ , along with constants  $a$  and  $b$ , then

$$\mathcal{L}^{-1}(a\widehat{U} + b\widehat{V}) = a\mathcal{L}^{-1}(\widehat{U}) + b\mathcal{L}^{-1}(\widehat{V}).$$

*Example:*  $U(t) = 1$  The Laplace transform is

$$\begin{aligned} \widehat{U}(s) &= \int_0^{\infty} e^{-st} dt \\ &= -\frac{1}{s} e^{-st} \Big|_{t=0}^{\infty}. \end{aligned}$$

This brings us to the question of what is the limiting value of  $e^{-st}$  as  $t \rightarrow \infty$ ? If  $\text{Re}(s) < 0$ , then the limit does not exist, while if  $s = 0$ , then  $1/s$  is not defined. Consequently, for  $\widehat{U}(s)$  to be defined it is required that  $\text{Re}(s) > 0$ . With this assumption, then  $\widehat{U}(s) = 1/s$ . ■

**Table 7.1** Inverse Laplace transforms

	$\widehat{U}(s) = \mathcal{L}(U)$	$\mathcal{L}^{-1}(\widehat{U})$
1.	$a\widehat{U}(s) + b\widehat{V}(s)$	$aU(t) + bV(t)$
2.	$\widehat{V}(s)\widehat{U}(s)$	$\int_0^t V(t-r)U(r)dr$
3.	$s\widehat{U}(s)$	$U'(t) + U(0)$
4.	$\frac{1}{s}\widehat{U}(s)$	$\int_0^t U(r)dr$
5.	$e^{-as}\widehat{U}(s)$	$U(t-a)H(t-a) \text{ for } a > 0$
6.	$\widehat{U}(s-a)$	$e^{at}U(t)$
7.	$\frac{1}{(s+a)^n}$	$\frac{1}{(n-1)!}t^{n-1}e^{-at} \text{ for } n = 1, 2, 3, \dots$
8.	$\frac{bs+c}{(s+a)^2+\omega^2}$	$e^{-at}\left(b\cos(\omega t) + \frac{c-ab}{\omega}\sin(\omega t)\right) \text{ for } \omega > 0$
9.	$\frac{cs+d}{(s+a)(s+b)}$	$\frac{1}{b-a}\left((bc-d)e^{-bt} - (ac-d)e^{-at}\right) \text{ for } a \neq b$
10.	$\frac{1}{\sqrt{s+a}}$	$\frac{1}{\sqrt{\pi t}}e^{-at}$
11.	$\frac{1}{(s+a)\sqrt{s+b}}$	$\frac{1}{\sqrt{b-a}}e^{-at}\operatorname{erf}\left(\sqrt{(b-a)t}\right) \text{ for } a \neq b$
12.	$\frac{1}{\sqrt{s}(\sqrt{s}+a)}$	$e^{a^2t}\operatorname{erfc}(a\sqrt{t})$
13.	$\frac{1}{s}e^{-as}$	$H(t-a) \text{ for } a > 0$
14.	$e^{-a\sqrt{s}}$	$\frac{a}{2\sqrt{\pi}}t^{-3/2}e^{-a^2/(4t)} \text{ for } a > 0$
15.	$\frac{1}{\sqrt{s}}e^{-a\sqrt{s}}$	$\frac{1}{\sqrt{\pi t}}e^{-a^2/(4t)} \text{ for } a > 0$
16.	$\frac{1}{s}e^{-a\sqrt{s}}$	$\operatorname{erfc}(a/(2\sqrt{t})) \text{ for } a > 0$
17.	$\frac{1}{s^{v+1}}e^{-a^2/(4s)}$	$\left(\frac{2}{a}\right)^v t^{v/2}J_v(a\sqrt{t}) \text{ for } \operatorname{Re}(v) > -1$
18.	$\frac{1}{q}e^{-cq} \text{ where } c > 0 \text{ and } q = \sqrt{(s+a)(s+b)}$	$e^{-(a+b)t/2}I_0[(a-b)\sqrt{t^2-c^2}/2]H(t-c)$

The Heaviside step function  $H(x)$  is defined in (7.20), the complementary error function  $\operatorname{erfc}(x)$  is given in (1.62), the error function  $\operatorname{erf}(x) = 1 - \operatorname{erfc}(x)$ , and  $J_v$ ,  $I_0$  are Bessel functions



*Example:*  $U(t) = e^{-t} \sin(3t)$  Using integration by parts, the Laplace transform is

$$\begin{aligned}\widehat{U}(s) &= \int_0^{\infty} \sin(3t) e^{-(s+1)t} dt \\ &= \left[ -\frac{s+1}{(s+1)^2 + 9} e^{-(s+1)t} \sin(3t) - \frac{3}{(s+1)^2 + 9} e^{-(s+1)t} \cos(3t) \right]_{t=0}^{\infty}.\end{aligned}$$

Assuming that  $\text{Re}(s) > -1$ , then

$$\widehat{U}(s) = -\frac{3}{(s+1)^2 + 9}. \quad \blacksquare$$

*Example*

$$U(t) = \begin{cases} 0 & \text{if } t \leq 1 \\ 2 & \text{if } 1 < t \leq 3 \\ -1 & \text{if } 3 < t \end{cases}$$

The Laplace transform is

$$\begin{aligned}\widehat{U}(s) &= \int_1^3 2e^{-st} dt - \int_3^{\infty} e^{-st} dt \\ &= -\frac{3}{s} e^{-3s} + \frac{2}{s} e^{-s}.\end{aligned}$$

It is interesting to see if we obtain the original function  $U(t)$  by taking the inverse transform of  $\widehat{U}(s)$ . First, from Property 1 from Table 7.1 it follows that

$$\mathcal{L}^{-1}(\widehat{U}) = -3\mathcal{L}^{-1}\left(\frac{1}{s}e^{-3s}\right) + 2\mathcal{L}^{-1}\left(\frac{1}{s}e^{-s}\right).$$

From Property 13 we then get that

$$\mathcal{L}^{-1}(\widehat{U}) = -3H(t-3) + 2H(t-1), \quad (7.19)$$

where  $H(x)$  is the *Heaviside step function*, and it is defined as

$$H(x) \equiv \begin{cases} 0 & \text{if } x < 0, \\ \frac{1}{2} & \text{if } x = 0, \\ 1 & \text{if } 0 < x. \end{cases} \quad (7.20)$$

Writing out the definition of  $H$  in (7.19), the inverse transform is

$$\mathcal{L}^{-1}(\widehat{U}) = \begin{cases} 0 & \text{if } t < 1, \\ 1 & \text{if } t = 1, \\ 2 & \text{if } 1 < t < 3, \\ \frac{1}{2} & \text{if } t = 3, \\ -1 & \text{if } 3 < t. \end{cases}$$

This result shows that  $\mathcal{L}^{-1}(\widehat{U}) = U$  at values of  $t$  where  $U$  is continuous, but at the jump points the inverse equals the average of the jump in the function. This is not unexpected given the requirement in (7.18). ■

*Example*

$$\widehat{U} = \frac{2}{s} - \frac{3}{s^2 + 4}$$

According to Property 7, from Table 7.1,  $\mathcal{L}^{-1}(\frac{1}{s}) = 1$ , and from Property 8,  $\mathcal{L}^{-1}((s^2 + 4)^{-1}) = \frac{1}{2} \sin(2t)$ . Using Property 1 it therefore follows that

$$\begin{aligned} U(t) &= \mathcal{L}^{-1}\left(\frac{2}{s} - \frac{3}{s^2 + 4}\right) \\ &= 2\mathcal{L}^{-1}\left(\frac{1}{s}\right) - 3\mathcal{L}^{-1}\left(\frac{1}{s^2 + 4}\right) \\ &= 2 - \frac{3}{2} \sin(2t). \quad \blacksquare \end{aligned}$$

### 7.1.2.1 Mathematical Requirements

Given the improper integral in (7.16), it is necessary to impose certain restrictions on the function  $U(t)$ , although the requirements are much less severe than for the Fourier transforms studied in Chap. 4. It is assumed that  $U(t)$  is piecewise continuous and has *exponential order*. This means that  $U$  grows no faster than a linear exponential function as  $t \rightarrow \infty$ . The specific requirement is that there is a constant  $\alpha$  so that

$$\lim_{t \rightarrow \infty} U e^{\alpha t} = 0. \quad (7.21)$$

As examples, any bounded function or any polynomial function has exponential order. On the other hand,  $e^{t^2}$  and  $e^{t^3}$  do not. With this, the Laplace transform (7.16) is defined for any  $s$  that satisfies  $\text{Re}(s) > \alpha$ , and this gives rise to what is known as the *half-plane of convergence* for the Laplace transform. This comes into play when

calculating the inverse transform (7.17), and the requirement is that  $a$  is in the half-plane of convergence. It is relatively easy to determine this half-plane from  $\widehat{U}$ . The requirement is that the half-plane of convergence is to the right of the singularities of  $\widehat{U}$  (see Fig. 7.4). As an example, if  $\widehat{U} = 1/s$ , then the half-plane of convergence is  $\text{Re}(s) > 0$ , while if  $\widehat{U} = 1/\sqrt{s(s-1)}$ , then the half-plane of convergence is  $\text{Re}(s) > 1$ .

One last comment to make before working out some of the properties of the Laplace transforms relates to the behavior of  $\widehat{U}$  when  $\text{Re}(s) \rightarrow \infty$ . Because of the negative exponential in the integral, it follows that

$$\lim_{\text{Re}(s) \rightarrow \infty} \widehat{U} = 0. \quad (7.22)$$

This limit assumes that the original function  $U$  is piecewise continuous and has exponential order. The reason this result is useful is that it can be used to help check for errors in a calculation. For example, if you find that  $\widehat{U} = s$ , or  $\widehat{U} = \sin(s)$ , or  $\widehat{U} = e^s$ , then an error has been made. The reason is that none of these functions satisfies (7.22).

### 7.1.2.2 Transformation of Derivatives

One of the hallmarks of the Laplace transform, as with most integral transforms, is that it converts differentiation into multiplication. To explain what this means, we use integration by parts to obtain the following:

$$\begin{aligned} \mathcal{L}(U') &= \int_0^\infty U' e^{-st} dt \\ &= U e^{-st} \Big|_{t=0}^\infty + s \int_0^\infty U e^{-st} dt \\ &= -U(0) + s\mathcal{L}(U). \end{aligned} \quad (7.23)$$

This formula can be used to find the transform of higher derivatives, and as an example

$$\begin{aligned} \mathcal{L}(U'') &= -U'(0) + s\mathcal{L}(U') \\ &= -U'(0) + s(-U(0) + s\mathcal{L}(U)) \\ &= s^2\mathcal{L}(U) - U'(0) - sU(0). \end{aligned} \quad (7.24)$$

Generalizing this to higher derivatives

$$\mathcal{L}(U^{(n)}) = s^n \mathcal{L}(U) - U^{(n-1)}(0) - sU^{(n-2)}(0) - \dots - s^{n-1}U(0). \quad (7.25)$$

### 7.1.2.3 Convolution Theorem

A common integral arising in viscoelasticity is a convolution integral of the form

$$T = \int_0^t G(t - \tau) V(\tau) d\tau. \quad (7.26)$$

Taking the Laplace transform of this equation we obtain

$$\begin{aligned} \mathcal{L}(T) &= \int_0^\infty \int_0^t G(t - \tau) V(\tau) e^{-st} d\tau dt \\ &= \int_0^\infty \int_\tau^\infty G(t - \tau) V(\tau) e^{-st} dt d\tau \\ &= \int_0^\infty \int_0^\infty G(r) V(\tau) e^{-s(r+\tau)} dr d\tau \\ &= \int_0^\infty V(\tau) e^{-s\tau} \left[ \int_0^\infty G(r) e^{-sr} dr \right] d\tau \\ &= \widehat{G}(s) \widehat{V}(s). \end{aligned}$$

Using the inverse transform this can be written as

$$\mathcal{L}^{-1}(\widehat{G}(s) \widehat{V}(s)) = \int_0^t G(t - \tau) V(\tau) d\tau. \quad (7.27)$$

This is Property 2, in Table 7.1, and it is known as the convolution theorem.

### 7.1.2.4 Solving the Problem for Linear Elasticity

The problem that will be solved using the Laplace transform consists of the wave equation

$$\frac{\partial^2 U}{\partial t^2} = c^2 \frac{\partial^2 U}{\partial X^2} + F(X, t), \quad (7.28)$$

where the boundary conditions are

$$U(0, t) = p(t), \quad U(\ell, t) = q(t), \quad (7.29)$$

and the initial conditions are

$$U(X, 0) = f(X), \quad \partial_t U(X, 0) = g(X). \quad (7.30)$$

It is understood that the only unknown is  $U(X, t)$ , and all the other functions in the above equations are given. The first step is to take the Laplace transform of both sides of the wave equation to obtain

$$\mathcal{L}(U_{tt}) = c^2 \mathcal{L}(U_{XX}) + \mathcal{L}(F). \quad (7.31)$$

Using (7.24), and the given initial conditions,

$$\mathcal{L}(U_{tt}) = s^2 \widehat{U} - g(X) - sf(X).$$

Also, because the transform is in the time variable,  $\mathcal{L}(U_{XX}) = \widehat{U}_{XX}$ . Introducing these observations into (7.31) we have that

$$c^2 \widehat{U}_{XX} - s^2 \widehat{U} = -\widehat{F}(X, s) - g(X) - sf(X). \quad (7.32)$$

The solution of this equation must satisfy the transform of the boundary conditions (7.29), and this means that

$$\widehat{U}(0, s) = \widehat{p}(s), \quad \widehat{U}(\ell, s) = \widehat{q}(s). \quad (7.33)$$

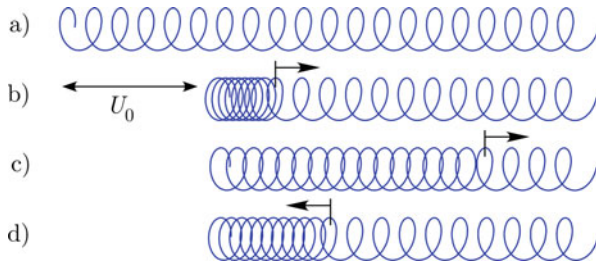
where  $\widehat{p} = \mathcal{L}(p)$  and  $\widehat{q} = \mathcal{L}(q)$ .

Solving (7.32) for  $\widehat{U}$  depends on what functions are used for the forcing, boundary, and initial conditions, and we consider two examples. Before doing this, note that by taking the Laplace transform that the initial conditions have become forcing functions in the differential equation (7.32). This limits the usefulness of this method. The reason is that even simple looking initial conditions can result in solutions of (7.32) that are complicated functions of the transform variable  $s$ . By complicated it is meant that the inverse transform is not evident, and even manipulating the contour integral in the definition of the inverse transform does not help. This observation should not be interpreted to mean that the method is a waste of time. Rather, it should be understood that the Laplace transform is an important tool for analyzing differential and integral equations, but like all other methods, it has limitations.

*Example (Slinky Wave)* Suppose there is no forcing, so  $F(X, t) = 0$ , and  $f(X) = g(X) = 0$ . Also, the boundary conditions are  $p = U_0$  and  $q = 0$ . Physically, this corresponds to holding an elastic bar at the right end, and then pushing on the left end a fixed amount  $U_0$ . A similar situation is shown in Fig. 7.5 for a slinky. What happens is that the disturbance propagates along the slinky, reaches the right end, reflects, and then moves leftward. The result is a disturbance that moves back and forth along the spring. This is mentioned as it is worth having some expectation on what the mathematics will produce.

Proceeding on to solving the problem, with the stated assumptions, (7.32) takes the form

$$c^2 \widehat{U}_{XX} - s^2 \widehat{U} = 0, \quad (7.34)$$



**Fig. 7.5** (a) A slightly extended slinky is held at  $X = 0$  and at  $X = \ell$ . (b) The left end is then moved a distance  $U_0$ , producing a compressed region in the spring. (c) This region spreads down the spring towards the right end. (d) When the compression reaches  $X = \ell$ , it reflects and then starts moving in the opposite direction. In an elastic spring this back-and-forth motion will continue indefinitely

and the boundary conditions (7.29) are

$$\widehat{U}(0, s) = \frac{U_0}{s}, \quad \widehat{U}(\ell, s) = 0. \quad (7.35)$$

This requires that  $\text{Re}(s) > 0$ . The general solution of (7.34) is

$$\widehat{U} = \alpha e^{sX/c} + \beta e^{-sX/c},$$

where  $\alpha$  and  $\beta$  are arbitrary constants. This function must satisfy the boundary conditions (7.35), and from this it follows that

$$\widehat{U}(X, s) = \frac{U_0}{s} \frac{\sinh(s(\ell - X)/c)}{\sinh(s\ell/c)}. \quad (7.36)$$

Now comes the big question, can we find the inverse transform of (7.36)? Some of the more extensive tables listing inverse Laplace transforms do include this particular function, but most do not. Given the propensity of second order differential equations to generate solutions involving the ratio of exponential functions, as in (7.36), it is worth deriving the inverse from scratch. The first step is to use the definition of the sinh function to write

$$\frac{\sinh(\alpha s)}{\sinh(\beta s)} = \frac{e^{\alpha s} - e^{-\alpha s}}{e^{\beta s} - e^{-\beta s}} = e^{-\beta s} \frac{e^{\alpha s} - e^{-\alpha s}}{1 - e^{-2\beta s}}.$$

It is assumed here that  $0 < \beta$ . Using the geometric series on the denominator,

$$\begin{aligned} \frac{\sinh(\alpha s)}{\sinh(\beta s)} &= e^{-\beta s} (e^{\alpha s} - e^{-\alpha s}) \left( 1 + e^{-2\beta s} + e^{-4\beta s} + \dots \right) \\ &= e^{-\beta s} (e^{\alpha s} - e^{-\alpha s}) + e^{-3\beta s} (e^{\alpha s} - e^{-\alpha s}) + e^{-5\beta s} (e^{\alpha s} - e^{-\alpha s}) + \dots \\ &= \sum_{n=1}^{\infty} e^{-(2n-1)\beta s} (e^{\alpha s} - e^{-\alpha s}). \end{aligned}$$

Now, using Property 13 from Table 7.1,

$$\begin{aligned}\mathcal{L}^{-1}\left(\frac{1}{s}e^{-bs}(e^{\alpha s} - e^{-\alpha s})\right) &= \mathcal{L}^{-1}\left(\frac{1}{s}e^{(\alpha-b)s}\right) - \mathcal{L}^{-1}\left(\frac{1}{s}e^{-(\alpha+b)s}\right) \\ &= H[t + (\alpha - b)] - H[t - (\alpha + b)].\end{aligned}$$

With this,

$$\mathcal{L}^{-1}\left(\frac{1}{s} \frac{\sinh(\alpha s)}{\sinh(\beta s)}\right) = \sum_{n=1}^{\infty} [H(t + \alpha - (2n-1)\beta) - H(t - \alpha - (2n-1)\beta)].$$

The inverse of (7.36) is, therefore,

$$U(X, t) = U_0 \sum_{n=1}^{\infty} [H(t + \kappa_{-n+1}) - H(t - \kappa_n)], \quad (7.37)$$

where

$$\kappa_n = \frac{1}{c}(-X + 2n\ell).$$

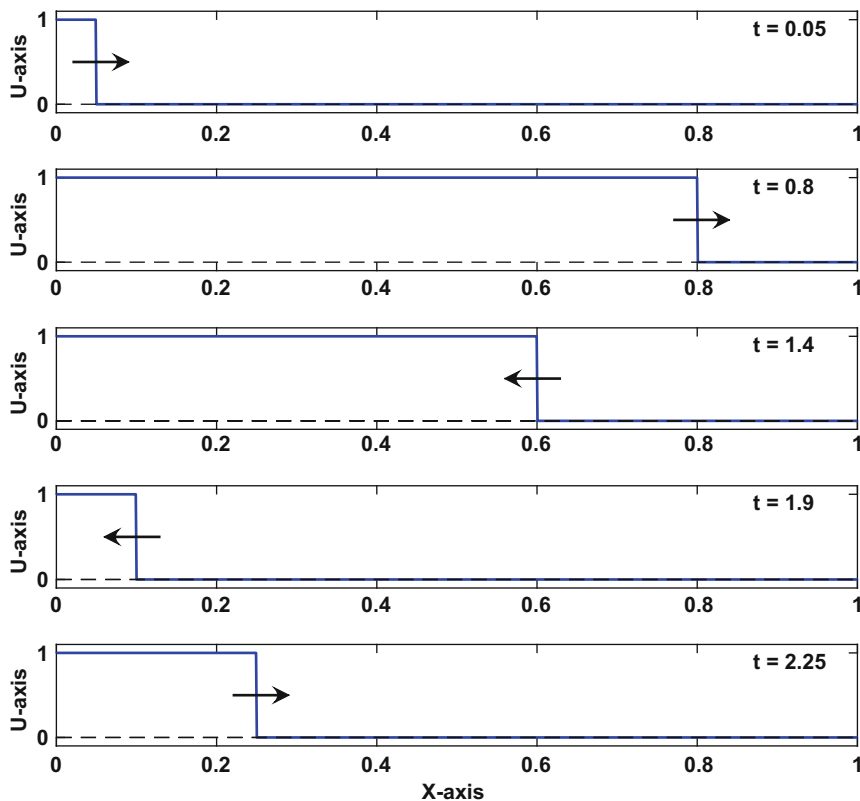
The solution is shown in Fig. 7.6, for  $\ell = c = U_0 = 1$ . As expected from the slinky analogy, the solution is a traveling wave that starts at  $X = 0$  and then moves back and forth over the bar. The amplitude is  $U_0 = 1$ , and the speed of the wave can be determined from the arguments of the Heaviside functions in (7.37). Namely, its speed is equal to  $c$ . This is not surprising as this is the speed of the traveling waves found using the method of characteristics, given in (7.13). As a final comment, there are different ways of writing the solution to this problem, and some are derived in Exercise 7.4. ■

*Example (Resonance)* In this example we investigate what happens to the bar when it is forced periodically. Specifically, it is assumed that the forcing function in (7.28) is  $F(X, t) = a(t) \cos(\kappa_n X)$ , where  $a(t) = \sin(\omega t)$ ,  $\kappa_n = n\pi/\ell$ , and  $n$  is a positive integer. The bar is assumed to be stress free at the ends, and so the boundary conditions are

$$\frac{\partial U}{\partial X} = 0 \text{ at } X = 0, \ell. \quad (7.38)$$

Taking the Laplace transform, the boundary conditions become

$$\frac{\partial \widehat{U}}{\partial X} = 0 \text{ at } X = 0, \ell. \quad (7.39)$$



**Fig. 7.6** Solution of the elastic bar given in (7.37). The solution consists of a traveling wave that starts at  $X = 0$  and then propagates back and forth along the  $X$ -axis

The initial conditions are  $f(X) = g(X) = 0$ . In this case (7.32) takes the form

$$c^2 \widehat{U}_{XX} - s^2 \widehat{U} = -\widehat{a}(s) \cos(\kappa_n X),$$

where  $\widehat{a} = \mathcal{L}(a)$ . The general solution of this equation is

$$\widehat{U} = \frac{\widehat{a}(s)}{\kappa_n^2 c^2 + s^2} \cos(\kappa_n X) + \alpha e^{sX/c} + \beta e^{-sX/c},$$

where  $\alpha$  and  $\beta$  are arbitrary constants. This solution must satisfy the boundary conditions (7.39), and from this it follows that

$$\widehat{U} = \frac{\widehat{a}(s)}{\kappa_n^2 c^2 + s^2} \cos(\kappa_n X). \quad (7.40)$$



Again, the big question, can we find the inverse transform of (7.40)? Using Property 8, with  $a = b = 0$  and  $\omega = \kappa_n c$ ,

$$\mathcal{L}^{-1}\left(\frac{1}{\kappa_n^2 c^2 + s^2}\right) = \frac{1}{\kappa_n c} \sin(\kappa_n c t).$$

Therefore, using the convolution property (7.27), it follows that

$$\begin{aligned} U(X, t) &= \mathcal{L}^{-1}\left(\frac{\widehat{a}(s)}{\kappa_n^2 c^2 + s^2} \cos(\kappa_n X)\right) \\ &= \cos(\kappa_n X) \mathcal{L}^{-1}\left(\widehat{a}(s) \frac{1}{\kappa_n^2 c^2 + s^2}\right) \\ &= \frac{1}{\kappa_n c} b(t) \cos(\kappa_n X), \end{aligned}$$

where

$$b(t) = \int_0^t a(t-r) \sin(\kappa_n c r) dr.$$

Given that  $a(t) = \sin(\omega t)$  it follows that

$$b(t) = \begin{cases} \frac{1}{\omega^2 - \kappa_n^2 c^2} (\omega \sin(\kappa_n c t) - \kappa_n c \sin(\omega t)) & \text{if } \omega \neq \kappa_n c, \\ -\frac{1}{2} t \cos(\kappa_n c t) + \frac{1}{2 \kappa_n c} \sin(\kappa_n c t) & \text{if } \omega = \kappa_n c. \end{cases} \quad (7.41)$$

This shows that when  $\omega \neq \kappa_n c$ , the displacement is a combination of two periodic functions. In contrast, when  $\omega = \kappa_n c$  the solution grows, becoming unbounded as  $t \rightarrow \infty$ . This is a phenomenon known as resonance, and it is a characteristic of linearly elastic systems.

The resonant frequencies are easily measured experimentally, and this provides a means to test the accuracy of the model. In the experiments of Bayon et al. (1993), an aluminum bar was tested and the first three measured resonant frequencies  $f_1$ ,  $f_2$ , and  $f_3$  are given in Table 7.2. Recall that circular and angular frequencies are related through the equation  $f = 2\pi\omega$ . In this case,  $\omega = \kappa_n c$  reduces to

$$f_n = \frac{n}{2\ell} \sqrt{\frac{E}{R_0}}. \quad (7.42)$$

To compare with the model, the bar in the experiment was 0.1647 m long. Also, using the conventional values for pure aluminum,  $E = 70758$  MPa and  $R_0 = 2700$  kg/m<sup>3</sup>. The resulting values of the angular frequencies are also shown in

**Table 7.2** Natural frequencies of an aluminum bar measured experimentally (Bayon et al. 1993), and computed using (7.42)

n	$f_n$ experimental	$f_n$ computed	Relative error
1	15,322 Hz	15,541 Hz	1.4%
2	30,644 Hz	31,082 Hz	1.4%
3	45,966 Hz	46,623 Hz	1.4%

Table 7.2. The rather small difference between the experimental and computed values is compelling evidence that the linear elastic model is appropriate here. ■

Given the need to be able to determine the material parameters in a model, the question comes up whether the measured values for  $f_n$  can be used to determine  $E$  and  $R_0$ . The best we can do with (7.42) is to determine the ratio  $E/R_0$ . How it might be possible to use the resonant frequencies to find the material and geometrical parameters is one of the core ideas in inverse problems. As an example, a classic paper in this area is, “Can you hear the shape of a drum,” by Kac (1966). Considerable work has been invested in solving inverse problems, and some of the more recent discoveries are discussed in Mueller and Siltanen (2012).

In a physical problem, the growth in the amplitude that occurs at the resonant frequency means that eventually the linear elasticity approximation no longer applies, and other effects come into play. These will generally mollify the amplitude, although not always. An example is the Tacoma Narrows Bridge. Although classic resonance was not the culprit, the same principle of unstable linear oscillations that feed large nonlinear motion was in play, and this eventually caused the bridge to collapse.

### 7.1.3 Geometric Linearity

The assumption in (7.1) that the stress is a linear function of the Lagrangian strain results in a linear momentum equation (7.2). This does not happen if any of the other strains listed in Table 6.4 are used. For example, when working in three dimensions it is conventional to use the Green strain  $\epsilon_g$ . From (6.50), the assumption that  $T = E\epsilon_g$  results in the momentum equation

$$\frac{\partial^2 U}{\partial t^2} = c^2 \left( 1 + \frac{\partial U}{\partial X} \right) \frac{\partial^2 U}{\partial X^2} + F.$$

In contrast to (7.2), this is a nonlinear wave equation for the displacement.

It is possible to obtain a linear momentum equation using the other strains, but it is necessary to impose certain restrictions on the motion. What is needed is geometric linearity, and to contrast this with our earlier assumption we have the following:

- *Material Linearity.* The assumption is that the stress-strain function is linear. Examples are  $T = E\epsilon$ ,  $T = E\epsilon_g$ , and  $\tau = E\epsilon_e$ . Linearity in this context is relative to a particular strain measure.
- *Geometric Linearity.* It is assumed that there are only small deformations in the sense that  $\epsilon \ll 1$ . This is often referred to as an assumption of infinitesimal deformations.

The assumption of geometric linearity means that the strains listed in Table 6.4 are, to first order, equal. For example,  $\epsilon_g = U_X + \frac{1}{2}U_X^2 \approx U_X = \epsilon$ . Similarly,  $\epsilon_e = u_x = U_X/(1 + U_X) \approx U_X = \epsilon$ . Also, recall that for an elastic material the constitutive assumption is that  $T = T(\epsilon)$ . Assuming  $T$  is smooth enough that Taylor's theorem can be used, then  $T'(\epsilon) = T'(0) + T''(0)\epsilon + \dots$ . Consequently, to first order, with the assumption of geometric linearity, the momentum equation for an elastic material reduces to the linear equation (7.2). This is assuming, of course, that  $E = T'(0) > 0$ .

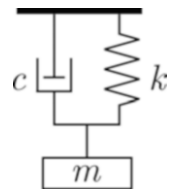
## 7.2 Viscoelasticity

The slinky example in the previous section is interesting but unrealistic from a physical point of view. The reason is that the traveling waves shown in Fig. 7.6 continue indefinitely. In contrast, in a real system the motion eventually comes to rest. One reason is that energy is lost due to dissipation. This is similar to what occurs when dropping an object and letting it fall through the air. The faster the object moves the greater the air resistance on the object. The usual assumption in this case is that there is a resistance force that is proportional to the object's velocity. This same idea is used when formulating the equations for a damped oscillator, and in the next section we use this observation to develop the theory of viscoelasticity.

### 7.2.1 Mass, Spring, Dashpot Systems

It is informative to review the equation for a damped oscillator, as shown in Fig. 7.7. From Newton's second law, the displacement  $u(t)$  of the mass in the mass, spring, dashpot system satisfies

**Fig. 7.7** Mass, spring, dashpot system



$$mu'' = F_s + F_d, \quad (7.43)$$

where  $m$  is the mass,  $F_s$  is the restoring force in the spring, and  $F_d$  is the damping force. Assuming the spring is linear, then from Hooke's law

$$F_s = -ku, \quad (7.44)$$

where  $k$  is a positive constant. The mechanism commonly used to produce damping involves a dashpot, where the resisting force is proportional to velocity. The associated constitutive assumption is

$$F_d = -cu', \quad (7.45)$$

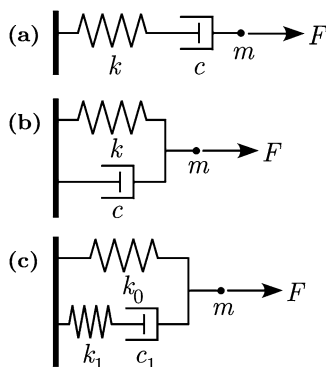
where  $c$  is a positive constant. With this, the total force  $F = F_s + F_d$ .

This example contains several ideas that will be expanded on below. First, it shows that the force includes an elastic component, which depends on displacement, and a damping component that depends on the velocity. As we saw earlier, when using a spring-mass system to help formulate a constitutive law for the stress, displacement is replaced with strain. Therefore, instead of assuming the force depends on displacement and velocity, in the continuum formulation the stress is assumed to depend on the strain  $\epsilon = \partial_X U$  and strain rate  $\epsilon_t = \partial_t(\partial_X U)$ . The question is, as always, exactly what function should we select. To help answer this question, we will examine spring and dashpot systems.

It is possible to generalize the above example and introduce the basic laws of viscoelasticity. This is done by putting the spring and dashpot in various configurations, and three of the more well studied are shown in Fig. 7.8.

We start with the series orientation shown in Fig. 7.8a. The point  $m$  moves due to a force  $F(t)$ , and its displacement  $u(t)$  equals the sum of the displacement  $u_s(t)$  of the spring and the displacement  $u_d(t)$  of the dashpot. Converting to velocities we have that  $u' = u'_s + u'_d$ . Now, according to Newton's third law, the force in the

**Fig. 7.8** Spring, dashpot systems used to derive viscoelastic models: (a) Maxwell element; (b) Kelvin-Voigt element; and (c) standard linear element



spring and dashpot equals  $-F$ . From (7.44) we get  $u_s = F/k$  and from (7.45) we have that  $u'_d = F/c$ . With this we obtain the following force, deflection relationship

$$u' = F'/k + F/c. \quad (7.46)$$

This gives rise to what is known as the Maxwell element in viscoelasticity.

In the next configuration, shown in Fig. 7.8b, the spring and dashpot are in parallel. For this we use the fact that forces add, and so  $F_s + F_d = -F$ . Also, the displacement of the spring and dashpot are the same, and both are equal to  $u(t)$ . With this we obtain

$$F = ku + cu'. \quad (7.47)$$

From this we get the Kelvin-Voigt element in viscoelasticity.

The third configuration, shown in Fig. 7.8c, gives rise to what is known as the standard linear element. The force in the upper spring is  $F_0 = -k_0u$ , while for the lower spring, dashpot the force satisfies  $u' = -F'_1/k_1 - F_1/c_1$ . The forces must balance, and this means that  $F = -F_0 - F_1$ . In this case,  $F_1 = -F - F_0$ , and so

$$\begin{aligned} u' &= -(-F - F_0)/k_1 - (-F - F_0)/c_1 \\ &= (F - k_0u)/k_1 + (F - k_0u)/c_1. \end{aligned}$$

Rearranging things, it follows that

$$F + a_1 F' = a_2 u + a_3 u', \quad (7.48)$$

where  $a_1 = c_1/k_1$ ,  $a_2 = k_0$  and  $a_3 = c_1(1+k_0/k_1)$ . The coefficients in this equation satisfy an inequality that is needed later. Because  $c_1 = k_1 a_1$ , then  $a_3 = a_1(k_1 + a_2)$ . With this we have that  $a_3 > a_1 a_2$ .

Each of the spring, dashpot examples can be generalized to a viscoelastic constitutive law that can be used in continuum mechanics. This is done by simply replacing  $u$  with the strain  $\epsilon$ ,  $u'$  with the strain rate  $\epsilon_t$ , and  $F$  with the stress  $T$ . After rearranging the constants in the formulas, the resulting viscoelastic constitutive laws are

$$\text{Maxwell model: } T + \tau_0 \frac{\partial T}{\partial t} = E \tau_1 \frac{\partial \epsilon}{\partial t}, \quad (7.49)$$

$$\text{Kelvin-Voigt model: } T = E \left( \epsilon + \tau_1 \frac{\partial \epsilon}{\partial t} \right), \quad (7.50)$$

$$\text{standard linear model: } T + \tau_0 \frac{\partial T}{\partial t} = E \left( \epsilon + \tau_1 \frac{\partial \epsilon}{\partial t} \right). \quad (7.51)$$

The strain in the above formulas, as usual, is

$$\epsilon = \frac{\partial U}{\partial X}.$$

In analogy with the linear elastic law (7.1), the constant  $E$  is the Young's modulus and it is assumed to be positive. The constants  $\tau_0$  and  $\tau_1$  have the dimensions of time, and are known as the dissipation time scales for the respective model. To be consistent with the expressions in (7.46)–(7.48),  $E$  and the  $\tau_i$ 's are assumed to be positive. In addition it is assumed in the standard linear model that  $\tau_0 < \tau_1$ . This condition comes from the same inequality that exists between the constants in (7.48).

## 7.2.2 Equations of Motion

The somewhat unusual forms of the viscoelastic constitutive laws generate several questions related to their mathematical and physical consequences. We begin with the mathematical questions, and with this in mind, remember that the reason for introducing a constitutive law is to complete the equations of motion. There are two functions that are solved for, which are the displacement and the stress. Using the standard linear model (7.51), and assuming there is no body force, the equations to solve are

$$R_0 \frac{\partial^2 U}{\partial t^2} = \frac{\partial T}{\partial X}, \quad (7.52)$$

$$T + \tau_0 \frac{\partial T}{\partial t} = E \left( \frac{\partial U}{\partial X} + \tau_1 \frac{\partial^2 U}{\partial X \partial t} \right). \quad (7.53)$$

To complete the problem, initial and boundary conditions must be specified and an example is presented below. Also, if one of the other viscoelastic models is used, then (7.53) would change accordingly.

*Example (Periodic Displacement)* A common testing procedure involves applying a periodic displacement to one end of the material, while keeping the other end fixed. Assuming the bar occupies the interval  $0 \leq X \leq \ell$ , then the associated boundary conditions are

$$U(0, t) = a \sin(\omega t), \text{ and } U(\ell, t) = 0. \quad (7.54)$$

We are going to solve the system of Eqs. (7.52) and (7.53). In doing so, it is assumed that the elastic modulus  $E$  and density  $R_0$  are known using one or more of the steady-state tests described in Sect. 6.7. Our goal here is to use the periodic displacement to determine the damping parameters  $\tau_0$  and  $\tau_1$ . This will be

accomplished by finding the periodic solution to the problem, which is the solution that appears long after the effects of the initial conditions have died out. To find this solution assume that

$$U(X, t) = \bar{U}(X)e^{i\omega t}, \quad (7.55)$$

and

$$T(X, t) = \bar{T}(X)e^{i\omega t}. \quad (7.56)$$

Using complex variables simplifies the calculations to follow, but it is necessary to rewrite the boundary condition at  $X = 0$  in (7.54) to fit this formulation. This will be done by generalizing it to

$$U(0, t) = ae^{i\omega t}. \quad (7.57)$$

It is understood that we are interested in the imaginary component of whatever expression we obtain. Now, substituting (7.55) and (7.56) into (7.53) we have that

$$\bar{T} = E \frac{1 + i\omega\tau_1}{1 + i\omega\tau_0} \frac{d\bar{U}}{dX}. \quad (7.58)$$

The momentum equation (7.52) in this case reduces to

$$\frac{d^2\bar{U}}{dX^2} = -\kappa^2 \bar{U},$$

where

$$\kappa^2 = \frac{R_0\omega^2}{E} \frac{1 + i\omega\tau_0}{1 + i\omega\tau_1}.$$

The general solution of this is  $\bar{U} = \alpha \exp(i\kappa X) + \beta \exp(-i\kappa X)$ . Imposing the two boundary conditions gives us the following solution:

$$\bar{U}(X) = a \frac{e^{i\kappa X} - e^{-i\kappa X + 2i\kappa\ell}}{1 - e^{2i\kappa\ell}}. \quad (7.59)$$

To simplify the analysis we will assume the bar is very long and let  $\ell \rightarrow \infty$ . With this in mind, note

$$\kappa^2 = \frac{R_0\omega^2}{E} \frac{1 + \omega^2\tau_0\tau_1 + i\omega(\tau_0 - \tau_1)}{1 + \omega^2\tau_1^2}.$$

Given that  $0 \leq \tau_0 < \tau_1$ , then  $\text{Re}(\kappa^2) > 0$  and  $\text{Im}(\kappa^2) < 0$ . From this we have that  $\text{Im}(\kappa) < 0$ , and so for large values of  $\ell$ , (7.59) reduces to

$$\overline{U}(X) = ae^{-i\kappa X}.$$

With this, the displacement is

$$U(X, t) = ae^{i(\omega t - \kappa X)}. \quad (7.60)$$

One of the reasons that experimentalists use this test is to compare the stress measured at  $X = 0$  with what is predicted from the model. With the solution in (7.60), and the formulas for the stress in (7.56) and (7.58), the stress at  $X = 0$  is

$$T(0, t) = -i\kappa aE \frac{1 + i\omega\tau_1}{1 + i\omega\tau_0} e^{i\omega t}. \quad (7.61)$$

To determine the imaginary component of this expression set

$$\begin{aligned} r_0 e^{i\delta} &= \kappa E \frac{1 + i\omega\tau_1}{1 + i\omega\tau_0} \\ &= \omega\sqrt{R_0 E} \sqrt{\frac{1 + i\omega\tau_1}{1 + i\omega\tau_0}} \\ &= \omega\sqrt{R_0 E} \sqrt{\frac{1 + \omega^2\tau_0\tau_1 + i\omega(\tau_1 - \tau_0)}{1 + \omega^2\tau_0^2}}. \end{aligned}$$

Taking the modulus of this we have that

$$r_0 = \omega\sqrt{R_0 E} \left( \frac{1 + \omega^2\tau_1^2}{1 + \omega^2\tau_0^2} \right)^{1/4}, \quad (7.62)$$

and taking the ratio of the imaginary and real components,

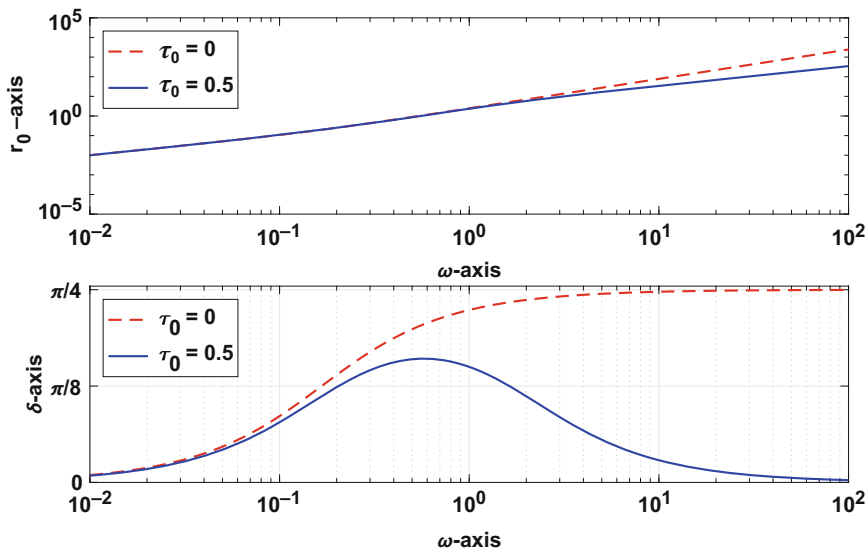
$$\tan(2\delta) = \frac{\omega(\tau_1 - \tau_0)}{1 + \omega^2\tau_0\tau_1}. \quad (7.63)$$

With this, (7.61) reduces to

$$T(0, t) = ar_0 \sin(\omega t + \delta - \pi/2). \quad (7.64)$$

We are now in a position to determine some of the effects of viscoelasticity. First, from (7.62), because  $1 + \omega^2\tau_1^2 > 1 + \omega^2\tau_0^2$ , the amplitude  $ar_0$  of the observed stress is increased due to the viscoelasticity. This conclusion is consistent with the



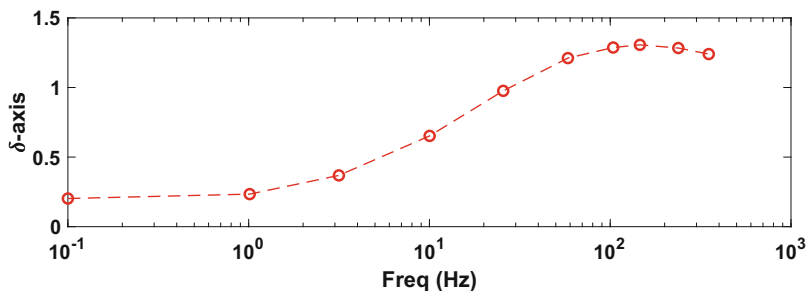


**Fig. 7.9** The amplitude  $r_0$  and phase  $\delta$  in response to a periodic forcing. Shown are the curves for a Kelvin-Voigt model,  $\tau_0 = 0$ , and for a standard linear model, where  $\tau_0 = 1/2$

understanding that damping increases the resistance to motion. However, as shown in Fig. 7.9, the  $r_0$  curves for the two viscoelastic models are rather similar, although they show some differences for very large values of  $\omega$ . What this means is that the  $r_0$  curve is not particularly useful in identifying which viscoelastic model to use. This is not the case with the phase  $\delta$ . As shown in (7.64), for a viscoelastic material the phase difference between the stress and displacement is  $\delta - \pi/2$ . The characteristics of  $\delta$  differ markedly between the two models. For the Kelvin-Voigt model, so  $\tau_0 = 0$ , the formula in (7.63) reduces to

$$\delta = \frac{1}{2} \arctan(\omega\tau_1). \quad (7.65)$$

In this case,  $\delta$  is a monotonically increasing function of  $\omega$ , and the larger the driving frequency the closer  $\delta$  gets to  $\pi/4$  (see Fig. 7.9). In comparison, for the standard linear model with  $0 < \tau_0 < \tau_1$ ,  $\delta$  reaches a maximum value when  $\omega = 1/\sqrt{\tau_0\tau_1}$ , and approaches zero as  $\omega \rightarrow \infty$ . This difference provides a simple test to determine which of the two models should be used. It is also useful for determining the damping parameters from experiment. If one is able to measure the frequency  $\omega_M$ , and phase  $\delta_M$ , for the maximum phase, then from (7.63) one finds that  $\tau_1 = [\tan(2\delta_M) + \sec(2\delta_M)]/\omega_M$  and  $\tau_0 = 1/(\tau_1\omega_M^2)$ . The derivation of this result is the subject of Exercise 7.17. To demonstrate that the frequency dependence shown in Fig. 7.9 does indeed occur in applications, data for porcine cartilage is



**Fig. 7.10** Measured values for  $\delta$  for porcine cartilage (Morita et al. 2002)

shown in Fig. 7.10. The dependence appears to follow the standard linear model. Also, note that cartilage is strongly viscoelastic. The reason is that if an elastic model is assumed, then  $\delta = 0$ , and this certainly does not happen in Fig. 7.10. ■

The previous example demonstrates how a mathematical model can be used in conjunction with experimental measurements to help test that the model is applicable, and to also determine some of the parameters. The focus of the inquiry was on the resulting stress at the end of the bar. It is also interesting to study the response within the bar. For example, with (7.60), the displacement has the form

$$U(X, t) = ae^{-\kappa_i X} \sin(\omega t - \kappa_r X), \quad (7.66)$$

where  $\kappa = \kappa_r - i\kappa_i$ . This is a traveling wave which has an amplitude that decays with  $X$ . A similar conclusion holds for the stress. Exactly how the viscoelasticity affects the properties of the wave is important in many applications, such as in geophysics when studying earthquakes, and this is explored in Exercise 7.14.

### 7.2.3 Integral Formulation

One of the attractive features of the Kelvin-Voigt model is that it provides an explicit formula for the stress. This can be substituted directly into the momentum equation (7.52), to produce a single equation for  $U$ , which avoids a system formulation as in (7.52) and (7.53). The other two viscoelastic models are implicit and require a solution of a differential equation to determine  $T$ . There are reasons, particularly when solving the problem numerically, why one would want to keep the problem in system form. However, there are also reasons why it is worth expressing the problem as a single equation.

To solve the Maxwell model (7.49), note that it is a linear first-order equation for  $T$ . Solving this equation one finds that

$$T(X, t) = T_0(X)e^{-t/\tau_0} + \int_0^t E \frac{\tau_1}{\tau_0} e^{(\tau-t)/\tau_0} \frac{\partial \epsilon}{\partial \tau} d\tau,$$

where  $T_0(X) = T(X, 0)$ . We will assume  $T(X, 0) = 0$ , so the above solution reduces to

$$T = \int_0^t G(t - \tau) \frac{\partial \epsilon}{\partial \tau} d\tau, \quad (7.67)$$

where

$$G(t) = E \frac{\tau_1}{\tau_0} e^{-t/\tau_0}. \quad (7.68)$$

Substituting this into (7.52) we obtain

$$R_0 \frac{\partial^2 U}{\partial t^2} = \int_0^t G(t - \tau) \frac{\partial^3 U}{\partial X^2 \partial \tau} d\tau, \quad (7.69)$$

which is an integro-differential equation for the displacement. The standard method for solving this equation is to use the Laplace transform. For the moment we will continue to concentrate on the formulation of the viscoelastic constitutive law and save the question of how to solve the problems until later.

The standard linear model is also a linear first-order equation for  $T$  that can be solved using an integrating factor. One finds that

$$T = \int_0^t G(t - \tau) \frac{\partial \epsilon}{\partial \tau} d\tau, \quad (7.70)$$

where

$$G(t) = E(1 + \kappa e^{-t/\tau_0}), \quad (7.71)$$

and  $\kappa = (\tau_1 - \tau_0)/\tau_0$  is a nonnegative constant. It has been assumed in deriving (7.70) that  $\epsilon = 0$  and  $T = 0$  at  $t = 0$ . With this we have obtained the same integral representation given in (7.67), except the function  $G$  is given in (7.71).

We now have two versions of the viscoelastic models. Those in (7.49)–(7.51) are differential equations and are examples of what are called rate-type laws. Expressing them in integral form we found that

$$T(X, t) = \int_0^t G(t - \tau) \frac{\partial \epsilon}{\partial \tau}(X, \tau) d\tau, \quad (7.72)$$

and this is known as a viscoelastic law of relaxation type. The function  $G$  is called the relaxation function. It is possible to rewrite the integral using integration by parts. The result is

$$T(X, t) = E\epsilon(X, t) + \int_0^t K(t - \tau)\epsilon(X, \tau)d\tau, \quad (7.73)$$

where  $K(t) = G'(t)$  and  $E = G(0)$ . Written this way, the stress is expressed as the sum of an elastic component and an integral associated with the damping in the system. Either version, (7.72) or (7.73), shows that the stress depends on the values of the strain, or strain rate, over the entire time interval. For this reason, the Maxwell and standard linear models apply to materials with memory. It might seem unreasonable to expect that the stress at the current time depends on what was happening a long time ago. However, the decaying exponential in (7.71) reduces the contribution from earlier times, and the smaller the dissipation time scale  $\tau_0$ , the less they contribute. In contrast, with the Kelvin-Voigt model the stress depends solely on the values of the strain and strain rate at the current time.

### 7.2.4 Generalized Relaxation Functions

The integral form of the stress law (7.72) is widely used in the engineering literature, and this is partly due to the information that is obtained from experiments. In many of the conventional tests used to determine material properties, the strain is imposed and the stress is measured. This information is then used to determine the parameters in the relaxation function. For this to work one must make a judicious choice for the functional form for  $G$ . The usual argument made in such situations is that real materials do not operate as a simple spring, dashpot system as in Fig. 7.8, but involve many such elements. The consequence of this observation is that one does not end up with one exponential, as in (7.68) and (7.71), but a relaxation function of the form

$$G(t) = E\left(1 + \sum \kappa_i e^{-t/\tau_i}\right). \quad (7.74)$$

An example spring, dashpot system that produce a multi-exponential relaxation function is shown in Fig. 7.14, and the specifics are worked out in Exercise 7.11. Although (7.74) is considered an improvement over the earlier simpler models, it still has flaws. Again, the argument is that because of the complexity of real materials, a finite number of elements is inadequate and one should use a continuous distribution. What happens in this case is that the sum in (7.74) is replaced with an integral. The resulting constitutive law for the relaxation function is

$$G(t) = E\left(1 + \int_0^\infty g(\tau)e^{-t/\tau}d\tau\right), \quad (7.75)$$

where  $g(\tau)$  is a nonnegative function. This transfers the question of how to pick  $G$  to what to take for  $g$ , which is not much of an improvement in terms of difficulty. The answer depends on the application. One approach is to attempt to formulate a general law that is still simple enough to allow analysis of the problem. For example, a commonly made choice used to model the viscoelastic properties of biological materials is

$$G(t) = E \left( 1 + \kappa \int_{\tau_1}^{\tau_2} \frac{1}{\tau} e^{-t/\tau} d\tau \right), \quad (7.76)$$

where  $\kappa$ ,  $\tau_1$ ,  $\tau_2$  are positive constants. This is known as the Neuberger-Fung relaxation function. Exactly how to determine the three constants from experiment is discussed in Fung (1993).

We are in a quagmire that is common in viscoelasticity, which is having multiple constitutive laws to pick from but not knowing exactly which one to use. The answer, again, depends on the application. To illustrate, let's reconsider the slinky example of the previous section. As noted earlier, assuming the bar is linearly elastic means that the motion observed in Fig. 7.6 never slows down, much less stops, and this was the motivation for introducing a viscoelastic model in the first place. We will assume that the damping, or dashpot, mechanism only acts when the bar is moving, and when at rest the bar can be modeled as a linearly elastic material. In terms of the differential forms in (7.49)–(7.51), this means that when  $\epsilon_t = 0$  and  $T_t = 0$  the formula reduces to  $T = E\epsilon$ . This eliminates the Maxwell model from consideration. This observation is why (7.50) and (7.51) are referred to as viscoelastic solids, while (7.49) is called a viscoelastic fluid. This still leaves open the question of whether to use a Kelvin-Voigt or a standard linear model. The answer bridges the mathematical and experimental worlds. It is not uncommon for an applied mathematician to ask an experimentalist to run a specific test that corresponds to a problem that the mathematician is able to solve. It is also not uncommon for the experimentalist to reply that the testing equipment does not have the particular capability that is requested. What is necessary in such cases is for the two to work out an experimental procedure that can provide useful information for building and testing the model. One that has found wide use in viscoelasticity involves periodic loading, and this was considered in an earlier example. There are certainly other methods, and a review of the possibilities can be found in Lakes (2004).

### 7.2.5 Solving Viscoelastic Problems

One of the standard tools for reducing viscoelastic models, both rate and integral type, is the Laplace transform. There are a couple of reasons for this. One is that it converts differentiation into multiplication. The second reason is it can handle the convolution integrals that arise with the integral type viscoelastic laws.

*Example (Deriving the Relaxation Function)* For the standard linear model the stress  $T$  is related to the strain  $\epsilon$  through the equation

$$T + \tau_0 \frac{\partial T}{\partial t} = E \left( \epsilon + \tau_1 \frac{\partial \epsilon}{\partial t} \right).$$

To solve this for  $T$  take the Laplace transform of both sides to obtain

$$\mathcal{L}(T) + \tau_0 \mathcal{L}(T_t) = E(\mathcal{L}(\epsilon) + \tau_1 \mathcal{L}(\epsilon_t)).$$

It is assumed that  $T = 0$  and  $\epsilon = 0$  at  $t = 0$ . With this, and using Property 1 from Table 7.1, and (7.23), we obtain

$$\widehat{T} + \tau_0 s \widehat{T} = E(\widehat{\epsilon} + \tau_1 s \widehat{\epsilon}).$$

Solving for  $\widehat{T}$  yields

$$\widehat{T} = E \frac{1 + \tau_1 s}{1 + \tau_0 s} \widehat{\epsilon}.$$

We are going to take the inverse transform to find  $T$ . At first glance it might appear that the convolution theorem, Property 2 from Table 7.1, can be used to find the inverse of the right-hand side of the equation. However, the function multiplying  $\widehat{\epsilon}$  does not satisfy (7.21), and therefore there is no inverse transform for this function. It is possible to modify the equation to get this to work, and the trick is to write the equation as

$$\widehat{T} = E \frac{1 + \tau_1 s}{s(1 + \tau_0 s)} s \widehat{\epsilon}. \quad (7.77)$$

From (7.23),  $\mathcal{L}^{-1}(s \widehat{\epsilon}) = \epsilon_t$ , and from Property 9 from Table 7.1

$$\mathcal{L}^{-1} \left( \frac{1 + \tau_1 s}{s(1 + \tau_0 s)} \right) = 1 + \left( 1 - \frac{\tau_1}{\tau_0} \right) e^{-t/\tau_0}.$$

Applying the convolution theorem to (7.77), and then using integration by parts, the stress is

$$T = \int_0^t \left( 1 + \left( \frac{\tau_1}{\tau_0} - 1 \right) e^{-(t-r)/\tau_0} \right) \frac{\partial \epsilon}{\partial r} dr.$$

This result agrees with the solution given in (7.70) that was obtained using an integrating factor. For this particular problem the integrating factor method is easier to use, but its limitation is that it only works on first-order equations. The Laplace transform, however, also works on higher-order problems, and this is important

for studying more complex viscoelastic models, such as those investigated in Exercises 7.11 and 7.12. ■

*Example (Solving an Integro-Differential Equation)* Suppose the bar is modeled as a Maxwell viscoelastic material, and the integral form of the stress law (7.67) is used. As shown in (7.69), the momentum equation in this case is

$$R_0 \frac{\partial^2 U}{\partial t^2} = \int_0^t G(t - \tau) \frac{\partial^3 U}{\partial X^2 \partial \tau} d\tau, \quad (7.78)$$

where  $G(t)$  is given in (7.68). It is assumed the bar occupies the interval  $0 \leq X < \infty$ , and the associated boundary conditions are

$$\frac{\partial U}{\partial X}(0, t) = F(t), \quad (7.79)$$

and

$$\lim_{X \rightarrow \infty} U(X, t) = 0. \quad (7.80)$$

The initial conditions are  $U(X, 0) = \partial_t U(X, 0) = 0$ . Taking the Laplace transform of (7.78) we obtain

$$R_0 \mathcal{L}(U_{tt}) = \mathcal{L} \left( \int_0^t G(t - \tau) \frac{\partial^3 U}{\partial X^2 \partial \tau} d\tau \right). \quad (7.81)$$

Because of the initial conditions, from (7.24), we have that  $\mathcal{L}(U_{tt}) = s^2 \mathcal{L}(U)$ . Also, from the convolution theorem we know that

$$\mathcal{L} \left( \int_0^t G(t - \tau) V(\tau) d\tau \right) = \mathcal{L}(G) \mathcal{L}(V).$$

Consequently, (7.81) takes the form

$$R_0 s^2 \widehat{U} = \mathcal{L}(G) \mathcal{L} \left( \frac{\partial^3 U}{\partial X^2 \partial t} \right). \quad (7.82)$$

Now, basic integration gives us

$$\begin{aligned} \mathcal{L}(G) &= \int_0^\infty E \frac{\tau_1}{\tau_0} e^{-t/\tau_0} e^{-st} dt \\ &= E \frac{\tau_1}{1 + \tau_0 s}. \end{aligned}$$

Also,

$$\begin{aligned}\mathcal{L}\left(\frac{\partial^3 U}{\partial X^2 \partial t}\right) &= s \mathcal{L}\left(\frac{\partial^2 U}{\partial X^2}\right) \\ &= s \frac{\partial^2}{\partial X^2} \mathcal{L}(U).\end{aligned}$$

Introducing these into (7.82) we obtain

$$R_0 s^2 \widehat{U} = E \frac{\tau_1 s}{1 + \tau_0 s} \frac{\partial^2 \widehat{U}}{\partial X^2}.$$

The general solution of this second order differential equation is

$$\widehat{U} = \alpha e^{\omega X} + \beta e^{-\omega X}, \quad (7.83)$$

where

$$\omega = \sqrt{\frac{R_0 s(1 + \tau_0 s)}{E \tau_1}}. \quad (7.84)$$

To find  $\alpha$  and  $\beta$  we take the Laplace transform of the boundary conditions (7.79) and (7.80) to find that  $\widehat{U}_X(0, s) = \widehat{F}$  and  $\widehat{U}(X, s) \rightarrow 0$  as  $X \rightarrow \infty$ . In this case (7.83) reduces to

$$\widehat{U}(X, s) = -\frac{1}{\omega} \widehat{F}(s) e^{-\omega X}. \quad (7.85)$$

From Property 18, in Table 7.1 we have that

$$\mathcal{L}^{-1}\left(\frac{1}{\omega} e^{-\omega X}\right) = \kappa e^{-t/(2\tau_0)} I_0\left(\frac{1}{2\tau_0} \sqrt{t^2 - \lambda^2}\right) H(t - \lambda), \quad (7.86)$$

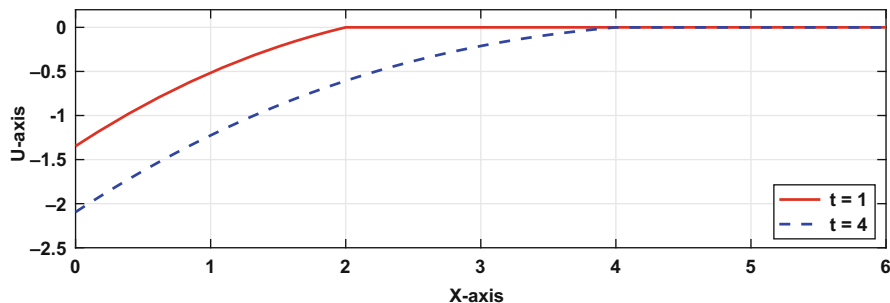
where  $\kappa = \sqrt{E \tau_1 / (R_0 \tau_0)}$  and  $\lambda = X/\kappa$ . In the above expression,  $I_0$  is the modified Bessel function of the first kind. With this, and the convolution theorem, it follows that

$$U(X, t) = H(t - \lambda) \int_0^{t-\lambda} Q(X, t - r) F(r) dr, \quad (7.87)$$

where

$$Q(X, t) = -\kappa e^{-t/(2\tau_0)} I_0\left(\frac{1}{2\tau_0} \sqrt{t^2 - \lambda^2}\right). \quad (7.88)$$





**Fig. 7.11** Solution of the Maxwell viscoelastic model as given in (7.87), in the case of when  $F(t) = \sin(t)$

An interesting conclusion that can be made is that the effects of the boundary condition move through the material with finite velocity. According to (7.87), the solution starts to be nonzero when  $t = \lambda$ , and the corresponding velocity is  $\sqrt{E\tau_1/(R_0\tau_0)}$ . On the other hand, the solution in (7.87) is not as satisfying a result as the previous example because the solution is in the form of a convolution integral involving a Bessel function. However, most of the math software programs, such as Maple and MATLAB, have the Bessel functions built in, so it is relatively easy to evaluate the integral. The result of such a calculation is shown in Fig. 7.11, which gives the solution at two time points. The finite velocity of the wave is clearly seen in this figure. ■

## Exercises

### Section 7.1

**7.1** A linearly elastic bar is stretched by applying a constant stress  $T_0$  to the right end. Assuming the original interval is  $0 \leq X \leq \ell_0$ , then the boundary conditions are  $U(0, t) = 0$  and  $T(\ell_0, t) = T_0$ . Assume there are no body forces.

- Find the steady-state solution for the density, displacement, and stress.
- What happens to the displacement and stress if Young's modulus is increased?

**7.2** The equations for the linearly elastic bar are given in (7.28)–(7.30). This exercise shows that not just any smooth function can be used in the displacement initial condition.

- Based on the impenetrability of matter requirement, what condition must be imposed on  $f(X)$  in (7.30)?
- Using the result from part (a), explain why it is not possible to take  $f(X) = 3X(\ell - X)/\ell$ , but it is possible to take  $f(X) = X(\ell - X)/(2\ell)$ .

**7.3** Find the solution of the problem for a linearly elastic bar with zero initial conditions, zero external forcing, and boundary conditions  $U(0, t) = 0$  and  $U(\ell, t) = U_0$ .

**7.4** This problem considers various ways to express the solution of the wave equation given in (7.37).

- (a) Show that  $-\kappa_{-n+1} \leq \kappa_n$ .  
 (b) Find a function  $I(x)$  so that it is possible to write (7.37), for  $0 \leq X < \ell$ , in the form

$$U(X, t) = U_0 \sum_{n=1}^{\infty} I[\alpha(t - t_n)],$$

where  $\alpha$  does not depend on  $n$ .

- (c) Show that (7.37) can be written as

$$U(X, t) = \begin{cases} U_0 & \text{if } 0 \leq X < q(t), \\ U_0/2 & \text{if } X = q(t), \\ 0 & \text{if } q(t) < X \leq \ell, \end{cases}$$

where  $q(t)$  is a  $2\ell/c$  periodic function.

**7.5** Solve the following problems by extending the method that was used in Sect. 7.1.1 to solve the wave equation.

- (a)

$$\frac{\partial^2 U}{\partial t^2} - 4 \frac{\partial^2 U}{\partial X^2} = 1,$$

where  $U(X, 0) = f(X)$  and  $\partial_t U(X, 0) = 0$ .

- (b)

$$\frac{\partial^2 U}{\partial t^2} + \frac{\partial U}{\partial t} = \frac{\partial^2 U}{\partial X^2} + \frac{\partial U}{\partial X},$$

where  $U(X, 0) = f(X)$  and  $\partial_t U(X, 0) = 0$ .

**7.6** A steel bar is forced periodically, and it is found that the first three resonant frequencies are 7861 Hz, 15,698 Hz, and 23,535 Hz (Bayon et al. 1994).

- (a) Explain why this result is consistent with the assumption that the bar is linearly elastic.  
 (b) If the bar is 0.32 mm long and has a density of 7893.16 kg/m<sup>3</sup>, find Young's modulus for the bar.  
 (c) The same experimentalists found that the wave speed in the bar is 5037 m/s. Use this to estimate the Young's modulus and compare the result from part (b).

## Section 7.2

**7.7** In elementary differential equations textbooks the equation of a mass, spring, dashpot system is stated to be  $mu'' + cu' + ku = 0$ , where  $u$  is the displacement of the mass from its equilibrium position. It is possible to find textbooks that use Fig. 7.8a to illustrate the system, and others that use Fig. 7.8b. Which is incorrect, and why?

**7.8** This problem considers how to rewrite a viscoelastic model as a single differential equation.

- (a) Show that the standard linear viscoelastic system (7.52) and (7.53) can be reduced to the single equation

$$\tau_0 R_0 \frac{\partial^3 U}{\partial t^3} + R_0 \frac{\partial^2 U}{\partial t^2} = E \frac{\partial^2 U}{\partial X^2} + \tau_1 E \frac{\partial^3 U}{\partial X^2 \partial t}.$$

- (b) Suppose that the initial conditions for (7.52) and (7.53) are  $U(X, 0) = f(X)$ ,  $\partial_t U(X, 0) = g(X)$ , and  $T(X, 0) = h(X)$ . The first two conditions can be used with the equation in part (a). Rewrite the third condition so it is a requirement only for  $U$  (and possibly its derivatives).
- (c) Show that the Maxwell system, which consists of (7.52) and (7.49), can be reduced to the single equation

$$\tau_0 R_0 \frac{\partial^2 U}{\partial t^2} + R_0 \frac{\partial U}{\partial t} = \tau_1 E \frac{\partial^2 U}{\partial X^2} + p(X),$$

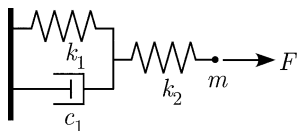
where  $p = R_0 g(X) + \tau_0 h'(X) - E \tau_1 f''(a)$ . The functions in the formula for  $p(X)$  come from the initial conditions:  $U(X, 0) = f(X)$ ,  $\partial_t U(X, 0) = g(X)$ , and  $T(X, 0) = h(X)$ .

**7.9** This problem concerns the system shown in Fig. 7.12, which is an example of what is known as a three-parameter viscoelastic solid.

- (a) Show that the force  $F$  and the displacement  $u$  satisfy

$$F + \frac{c_1}{k_1 + k_2} F' = \frac{k_1 k_2}{k_1 + k_2} u + \frac{c_1 k_2}{k_1 + k_2} u'.$$

**Fig. 7.12** Three-parameter viscoelastic solid studied in Exercise 7.9



- (b) Show that the continuum version of the result from part (a) has the form

$$T + \tau_0 \frac{\partial T}{\partial t} = E \left( \epsilon + \tau_1 \frac{\partial \epsilon}{\partial t} \right),$$

where  $0 < \tau_0 < \tau_1$ .

- (c) The resulting viscoelastic system consists of solving (7.52) along with the equation from part (b). Show that this can be rewritten as the single equation

$$\tau_0 R_0 \frac{\partial^3 U}{\partial t^3} + R_0 \frac{\partial^2 U}{\partial t^2} = E \frac{\partial^2 U}{\partial X^2} + \tau_1 E \frac{\partial^3 U}{\partial X^2 \partial t}.$$

- (d) Show that the viscoelastic constitutive law in part (b) can be expressed in integral form (7.67), where

$$G(t) = E(1 + \kappa e^{-t/\lambda}).$$

Assume that  $\epsilon = 0$  and  $T = 0$  at  $t = 0$ . Also, show that  $\kappa$  and  $\lambda$  are positive.

- (e) Discuss the similarities, and differences, between the results from (a) - (d) with the formulas for the standard linear model.

**7.10** This problem concerns the system shown in Fig. 7.13, which is an example of what is known as a three-parameter viscoelastic fluid.

- (a) Show that the force  $F$  and the displacement  $u$  satisfy

$$F + \frac{c_1 + c_2}{k_1} F' = c_2 u' + \frac{c_1 c_2}{k_1} u''.$$

- (b) Show that the continuum version of the result from part (a) has the form

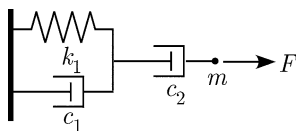
$$T + \tau_0 \frac{\partial T}{\partial t} = \tau_1 \frac{\partial \epsilon}{\partial t} + \tau_2 \frac{\partial^2 \epsilon}{\partial t^2},$$

where the  $\tau_i$ 's are positive with  $\tau_0 \tau_1 < \tau_2$ .

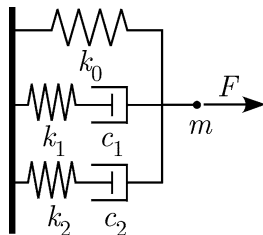
- (c) Show that the viscoelastic constitutive law in part (b) can be expressed in integral form as

$$T = \kappa_1 \epsilon' + \int_0^t G(t-s) \epsilon'(s) ds,$$

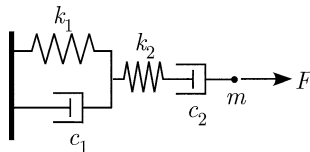
**Fig. 7.13** Three-parameter viscoelastic fluid studied in Exercise 7.10. This is known as the Jeffrey model



**Fig. 7.14** Five-parameter model for Exercise 7.11



**Fig. 7.15** The Burger viscoelastic model used in Exercise 7.12



where

$$G(t) = \kappa_2 e^{-t/\lambda}.$$

Assume that  $\epsilon = \epsilon' = 0$  and  $T = 0$  at  $t = 0$ . Also, show that  $\kappa_i$ 's and  $\lambda$  are positive.

**7.11** This problem concerns the five-parameter model shown in Fig. 7.14.

- Derive the differential equation that relates the force  $F$  with the displacement  $u$ .
- Show that the continuum version of the result from part (a) has the form

$$T + \tau_0 \frac{\partial T}{\partial t} + \tau_1 \frac{\partial^2 T}{\partial t^2} = E \left( \epsilon + \tau_2 \frac{\partial \epsilon}{\partial t} + \tau_3 \frac{\partial^2 \epsilon}{\partial t^2} \right),$$

where the  $\tau_i$ 's are positive, with  $\tau_0 < \tau_2$  and  $\tau_1 < \tau_3$ .

- Show that the viscoelastic constitutive law in part (b) can be expressed in integral form (7.67), where

$$G(t) = E(1 + \kappa_1 e^{-t/\lambda_1} + \kappa_2 e^{-t/\lambda_2}).$$

Assume that  $\epsilon = \epsilon' = 0$  and  $T = T' = 0$  at  $t = 0$ . Also, show that the  $\lambda_i$ 's are positive.

**7.12** This problem concerns the four-parameter model shown in Fig. 7.15, what is known as the Burger model.

- Derive the differential equation that relates the force  $F$  with the displacement  $u$ .

- (b) Show that the continuum version of the result from part (a) has the form

$$T + \tau_0 \frac{\partial T}{\partial t} + \tau_1 \frac{\partial^2 T}{\partial t^2} = \tau_2 \frac{\partial \epsilon}{\partial t} + \tau_3 \frac{\partial^2 \epsilon}{\partial t^2},$$

where the  $\tau_i$ 's are positive, with  $4\tau_1 < \tau_0^2$ .

- (c) Show that the viscoelastic constitutive law in part (b) can be expressed in integral form (7.67), where

$$G(t) = \kappa_1 e^{-t/\lambda_1} + \kappa_2 e^{-t/\lambda_2}.$$

Assume that  $\epsilon = \epsilon' = 0$  and  $T = T' = 0$  at  $t = 0$ . Also, show that the  $\lambda_i$ 's are positive.

**7.13** In some applications it is easier to work with the stress rather than the displacement. This problem investigates this for the standard linear model.

- (a) Derive (7.48).  
 (b) Starting from (7.51), show that

$$\epsilon = \int_0^t J(t - \tau) \frac{\partial T}{\partial \tau} d\tau.$$

Assume here that  $\epsilon = 0$  and  $T = 0$  at  $t = 0$ . The function  $J$  is called the creep function.

- (c) Show that

$$\epsilon = J(0)T + \int_0^t J'(t - \tau)T d\tau.$$

- (d) Use the result in part (b) to transform (7.53) into an equation for the stress  $T$ .  
 (e) By taking the Laplace transform of the creep and relaxation forms of the constitutive laws show that

$$\int_0^t G(s)J(t - s)ds = t.$$

**7.14** This problem investigates the traveling waves that are obtained in the periodic displacement example using the standard linear model.

- (a) Assuming  $\ell$  is large, and letting  $\kappa = \kappa_r - i\kappa_i$ , show that (7.55) and (7.59) reduce to

$$U(X, t) = ae^{-\kappa_i X} \sin(\omega t - X\kappa_r).$$

Find the corresponding expression for the stress  $T(X, t)$ .

(b) Show that for high frequencies

$$\kappa \sim \sqrt{\frac{R_0 \tau_0}{E \tau_1}} \omega \left( 1 - i \frac{\tau_1 - \tau_0}{2 \tau_0 \tau_1 \omega} + O\left(\frac{1}{\omega^2}\right) \right).$$

(c) Show that for low frequencies

$$\kappa \sim \sqrt{\frac{R_0}{E}} \omega \left( 1 - i \frac{1}{2} (\tau_1 - \tau_0) \omega + O(\omega^2) \right).$$

(d) Suppose the elastic modulus  $E$  and density  $R_0$  are known. Can the phase velocity  $v_p = \omega/\kappa_r$  of the wave, measured at both low and high frequencies, be used to determine the two viscoelastic constants  $\tau_0$  and  $\tau_1$ ? If the amplitude  $ae^{-\kappa_i X}$  of the wave is also measured at both low and high frequencies, does this help in determining the two viscoelastic constants  $\tau_0$  and  $\tau_1$ ?

**7.15** One of the consequences of damping is that it can mollify the effects of resonance. As an example, suppose that in (7.59) the viscoelasticity is turned off by letting  $\tau_0 = \tau_1 = 0$ . In this case there are frequencies for which (7.59) is undefined. Relate these to the resonance frequencies found in (7.41). Explain why (7.59) does not have this particular problem when the viscoelasticity is turned on (remember that  $\tau_0 < \tau_1$ ).

**7.16** This problem investigates the differences in the viscoelastic models when a periodic forcing is used. The boundary conditions in this case are

$$T(0, t) = b \sin(\omega t), \quad U(\ell, t) = 0.$$

Assume the standard linear viscoelastic model is used.

- Assuming a periodic solution, find  $\bar{U}(X)$  and  $\bar{T}(X)$ .
- Find  $\bar{U}(0)$  assuming  $\ell \rightarrow \infty$ . Your answer should be in terms of one trig function, similar to what was done for the stress in (7.64).
- In the experiments one measures the displacement at the end and compares the data with the predictions from the model. One objective is to determine the viscoelastic parameters in the model. Does the periodic stress boundary condition provide any information not learned from the periodic displacement boundary condition?

**7.17** This problem explores some of the consequences of the periodic displacement example of Sect. 7.2.2. Assume that  $0 < \tau_0 < \tau_1$ .

- Assume that the maximum value of  $\delta$ , as determined from (7.63), is  $\delta_M$  and occurs at frequency  $\omega_M$ . Show that  $\tau_1 = [\tan(2\delta_M) + \sec(2\delta_M)]/\omega_M$  and  $\tau_0 = 1/(\tau_1 \omega_M^2)$ .
- Use your results from part (a) to estimate  $\tau_1$  that was used in Fig. 7.9.

### ***Additional Questions on the Laplace Transform***

**7.18** Find the Laplace transform of the following functions. Make sure to state if there are conditions on  $s$ .

- (a)  $f(t) = te^{\alpha t}$
- (b)  $f(t) = \cosh^2 t$
- (c)  $Si(t) = \int_0^t \frac{\sin(r)}{r} dr$
- (d)  $f(t) = \frac{1}{t} \sin(t)$
- (e)  $f(t) = \sqrt{t}$

**7.19** Using the Laplace transform, solve  $y'' + 4y = f(t)$ , where  $y(0) = 0$ ,  $y'(0) = -1$ , and

$$f(t) = \begin{cases} \cos(2t) & \text{if } 0 \leq t < \pi, \\ 0 & \text{otherwise.} \end{cases}$$

**7.20** Using the Laplace transform, solve the system of equations

$$\begin{aligned} x' &= 3x - 4y \\ y' &= 2x + 3y, \end{aligned}$$

where  $x(0) = 1$  and  $y(0) = 0$ .

**7.21** This problem concerns solving the diffusion equation

$$Du_{xx} = u_t, \quad \text{for } \begin{cases} 0 < x < \infty, \\ 0 < t, \end{cases}$$

where  $u(x, 0) = 0$ ,  $u \rightarrow 0$  as  $x \rightarrow \infty$ , and

$$u(0, t) = \begin{cases} T & \text{if } 0 < t \leq b, \\ 0 & \text{if } b < t. \end{cases}$$

Using the Laplace transform, find the solution of this problem.

**7.22** Using the Laplace transform solve the integral equation

$$u(t) - \int_0^t e^{t-r} u(r) dr = f(t).$$

Assume this holds for  $0 \leq t$  and that  $f(t)$  is continuous.

**7.23** In solving the tautochrone problem one finds that it is necessary to solve the integral equation

$$\int_0^t \frac{u(r)}{\sqrt{t-r}} dr = \alpha, \quad \text{for } 0 < t,$$



where  $\alpha$  is a positive constant. Find the solution using the Laplace transform.

**7.24** Using the Laplace transform solve the integral equation

$$u(t) + \int_0^t \frac{u(r)}{\sqrt{t-r}} dr = f(t),$$

where  $f(t)$  is smooth and satisfies  $f(0) = 0$ .

### Additional Questions

**7.25** This problem explores the effect on the solution when using different materially linear theories. The two constitutive laws that are compared are: (i)  $T = EU_X$ , and (ii)  $\tau = Eu_x$ . As usual, let  $\epsilon = U_X$  and  $\epsilon_e = u_x$

- Transform constitutive law (ii) into material coordinates, that is, transform it into an expression involving  $T$  and  $\epsilon$ . For labeling purposes, identify this stress as  $T_{ii}$  and label the one from (i) as  $T_i$ . On the same axes, sketch  $T_{ii}$  and  $T_i$  for  $-1 < \epsilon < \infty$ .
- Transform constitutive law (i) into spatial coordinates, that is, transform it into an expression involving  $\tau$  and  $\epsilon_e$ . For labeling purposes, identify this stress as  $\tau_i$  and label the one from (ii) as  $\tau_{ii}$ . On the same axes, sketch  $\tau_i$  and  $\tau_{ii}$  for  $-\infty < \epsilon_e < 1$ .
- Show that  $T_{ii} < T_i$  if  $\epsilon \neq 0$ .
- Show that  $\tau_{ii} > \tau_i$  if  $\epsilon_e \neq 0$ .
- Suppose the stress in the bar becomes unbounded for large tensile strains. Is  $T_{ii}$  or  $T_i$  the more appropriate constitutive law?
- Suppose the stress in the bar becomes unbounded for large compressive strains. Is  $T_{ii}$  or  $T_i$  the more appropriate constitutive law?

**7.26** This problem explores what happens with a viscoelastic constitutive law when there is a jump in the solution. To do this, assume that at a given position  $X$ , the stress and strain are smooth except for a jump discontinuity when  $t = t_s$ .

- By integrating the constitutive law (7.51) over the time interval  $t_s - \Delta t \leq t \leq t_s + \Delta t$ , show that an expression of the following form is obtained,

$$\begin{aligned} & \tau_0 [T(X, t_s + \Delta t) - T(X, t_s - \Delta t)] \\ &= E\tau_1 [\epsilon(X, t_s + \Delta t) - \epsilon(X, t_s - \Delta t)] + \int_{t_s - \Delta t}^{t_s + \Delta t} q(X, t) dt. \end{aligned}$$

- By letting  $\Delta t \rightarrow 0$ , show that

$$\tau_0 [T(X, t_s^+) - T(X, t_s^-)] = E\tau_1 [\epsilon(X, t_s^+) - \epsilon(X, t_s^-)].$$

This states how the stress and strain behave across a jump, similar to what is obtained from the Rankine-Hugoniot condition for traffic flow.

- (c) A common experiment is to apply a constant stress at one end of the bar, which is assumed here to be at  $X = 0$ . This produces what is known as a creep response, and the associated boundary condition is  $T(0, t) = T_0$  for  $t > 0$ . Assume that for  $t < 0$  the bar is at rest with  $T = 0$  and  $\epsilon = 0$ . Using the standard linear model show that an expression of the following form is obtained

$$E_i \frac{\partial U}{\partial X}(0, 0^+) = T_0,$$

where  $E_i = E\tau_1/\tau_0$ . In engineering  $E_i$  is called the instantaneous elastic modulus. Explain why it is larger than the elastic modulus.

- (d) Find the instantaneous elastic modulus when using the Maxwell model.

**7.27** The Oldroyd model for a viscoelastic fluid assumes a constitutive law for the stress, in spatial coordinates, of the form

$$\sigma + \tau_1 \frac{\partial_a \sigma}{\partial t} = 2\mu_0 \left( d + \tau_2 \frac{\partial_a d}{\partial t} \right).$$

In this equation, the odd looking time derivative is called a convected derivative. Given a function  $p(x, t)$ , it is defined as

$$\frac{\partial_a p}{\partial t} \equiv \frac{Dp}{Dt} - 2adp,$$

where  $a$  is a constant that satisfies  $-1 \leq a \leq 1$ . Also,  $d = \partial_x v$ , where  $v(x, t)$  is the velocity, is known as the rate of deformation function. Note that we are considering one-dimensional motion, as assumed in this and the previous chapter.

- (a) The above constitutive law looks similar to the standard linear viscoelastic model (7.51), except it is expressed in spatial coordinates and it uses the convected derivative. Rewrite the law in material coordinates, so the result is an equation in terms of  $T$  and  $\epsilon$ .
- (b) An Oldroyd-A fluid is obtained when  $a = -1$ , which corresponds to what is called the lower convective derivative. Show that, in this case, your constitutive law in part (a) reduces to

$$T + \tau_1 \left( \partial_t T + 2TZ \right) = 2\mu_0 \left( Z + \tau_2 \partial_t Z + 2\tau_2 Z^2 \right),$$

where  $Z = \partial_t \ln(1 + \epsilon)$ .

- (c) An Oldroyd-B fluid is obtained when  $a = 1$ , which corresponds to what is called the upper convective derivative. Show that, in this case, your constitutive law in part (a) reduces to

$$T + \tau_1 \left( \partial_t T - 2TZ \right) = 2\mu_0 \left( Z + \tau_2 \partial_t Z - 2\tau_2 Z^2 \right),$$

where  $Z = \partial_t \ln(1 + \epsilon)$  is the Hencky strain rate.

# Chapter 8

## Continuum Mechanics: Three Spatial Dimensions



### 8.1 Introduction

The water in the ocean, the air in the room, and a rubber ball have a common characteristic, they appear to completely occupy their respective domains. What this means is that the material occupies every point in the domain. This observation is the basis of the continuum approximation, and it was used in Sect. 5.2 to define continuum variables such as density and flux. These variables can be usefully defined as long as the individual nature of the constituent particles are not apparent. So, for example, the continuum approximation is not appropriate on the nanometer scale, because atomic radii range from 0.2 to 3.0 nm. It can, however, be used down to the micron level. As an example, at 15 °C, and one atmosphere, there are approximately  $3 \times 10^7$  molecules in a cubic micron of air. Similarly, for water at room temperature there are approximately  $3 \times 10^{10}$  molecules in a cubic micron, and for a metal such as copper there are approximately  $10^{11}$  atoms in a cubic micron. Consequently, the averaging on which the continuum approximation is based is applicable down to the micron scale. This is why continuum models are commonly used for micro-devices, which involve both electrical and mechanical components. At the other extreme, continuum models are used to study the motion of disk galaxies and, more recently, to investigate the existence and properties of the “dark fluid” proposed to be responsible for the expansion of the universe. This range of applicability is why the continuum approximation, and the subsequent equations of motion, play a fundamental role in most branches of science and engineering. From a mathematical standpoint, the problems that come from continuum models have been almost single-handedly responsible for the development of an area central to applied mathematics, and this is the subject of nonlinear partial differential equations.

In this chapter the fundamental concepts of continuum mechanics are introduced, and they are then used to derive equations of motion for viscous fluids and elastic solids.

## 8.2 Material and Spatial Coordinates

To define the material coordinate system, assume that at  $t = 0$  a particular point in the material is located at  $\mathbf{x} = \mathbf{X}$ . It is assumed that as the material moves, the position of the point is given as  $\mathbf{x} = \mathcal{X}(\mathbf{X}, t)$ . To be consistent, the position function must satisfy  $\mathcal{X}(\mathbf{X}, 0) = \mathbf{X}$ . The resulting displacement and velocity functions are defined as

$$\mathbf{U}(\mathbf{X}, t) \equiv \mathcal{X}(\mathbf{X}, t) - \mathbf{X}, \quad (8.1)$$

and

$$\mathbf{V}(\mathbf{X}, t) \equiv \frac{\partial \mathbf{U}}{\partial t}. \quad (8.2)$$

Because  $\mathcal{X}(\mathbf{X}, 0) = \mathbf{X}$ , it follows that  $\mathbf{U}(\mathbf{X}, 0) = \mathbf{0}$ .

Instead of following particles as they move, one can select a spatial location and then let them come to you. This is the viewpoint taken for spatial coordinates. In this system, the displacement function is denoted as  $\mathbf{u}(\mathbf{x}, t)$ , and the velocity is  $\mathbf{v}(\mathbf{x}, t)$ . As is usual for displacement functions, it is required that  $\mathbf{u}(\mathbf{x}, 0) = \mathbf{0}$ .

*Example* Suppose a particle that started at location  $(1, -1, 1)$  is, at  $t = 2$ , located at  $(3, 0, -1)$ .

*Material Coordinates:* For this particle,  $\mathbf{X} = (1, -1, 1)$ , and its displacement at  $t = 2$  is  $(3, 0, -1) - (1, -1, 1) = (2, 1, -2)$ . In other words,

$$\mathbf{U}(\mathbf{X}, 2) = (2, 1, -2), \quad \text{for } \mathbf{X} = (1, -1, 1). \quad (8.3)$$

We also have that

$$\mathcal{X}(\mathbf{X}, 0) = (1, -1, 1), \quad \text{for } \mathbf{X} = (1, -1, 1),$$

and

$$\mathcal{X}(\mathbf{X}, 2) = (3, 0, -1), \quad \text{for } \mathbf{X} = (1, -1, 1).$$

*Spatial Coordinates:* At  $t = 2$ , the displacement of the particle located at  $(3, 0, -1)$  is  $(2, 1, -2)$ . In other words,

$$\mathbf{u}(\mathbf{x}, 2) = (2, 1, -2), \quad \text{for } \mathbf{x} = (3, 0, -1). \quad \blacksquare \quad (8.4)$$

The spatial system is the one usually used for fluids, which includes both gases and liquids. As an example, when measuring the properties of the atmosphere, observers are often fixed, and not moving with the air. This is the viewpoint taken by the spatial coordinate system, and hence the reason why it is the default system in

fluid dynamics. In contrast, the material system is associated with solid mechanics. One reason is that the configuration at  $t = 0$ , what is known as the reference state, is usually known for a solid. The fact is, however, that some of the more interesting contemporary applications of continuum mechanics involve both fluid and solid components. A particularly rich area for this is biology, which includes the study of how birds fly and the study of the internal mechanisms of cell function. For this reason, both coordinate systems need to be understood, and both are studied in this chapter.

The material and spatial descriptions for the displacement and velocity functions must be consistent. This was demonstrated in the last example, as given in (8.3) and (8.4). To express this in a general form, for a particle with position function  $\mathbf{x} = \mathcal{X}(\mathbf{X}, t)$  it is required that  $\mathbf{U}(\mathbf{X}, t) = \mathbf{u}(\mathbf{x}, t)$ , and  $\mathbf{V}(\mathbf{X}, t) = \mathbf{v}(\mathbf{x}, t)$ . More expansively, the required consistency conditions can be written as

$$\mathbf{U}(\mathbf{X}, t) = \mathbf{u}(\mathcal{X}(\mathbf{X}, t), t), \quad (8.5)$$

and

$$\mathbf{V}(\mathbf{X}, t) = \mathbf{v}(\mathcal{X}(\mathbf{X}, t), t). \quad (8.6)$$

The transformation between the two coordinate systems is assumed to be invertible, and so it is possible to solve  $\mathbf{x} = \mathcal{X}(\mathbf{X}, t)$  uniquely for  $\mathbf{X}$ . Writing the solution as  $\mathbf{X} = \mathbf{x}(\mathbf{x}, t)$ , then

$$\mathbf{u}(\mathbf{x}, t) = \mathbf{U}(\mathbf{x}(\mathbf{x}, t), t), \quad (8.7)$$

and

$$\mathbf{v}(\mathbf{x}, t) = \mathbf{V}(\mathbf{x}(\mathbf{x}, t), t). \quad (8.8)$$

The above formulas will be invaluable when converting the equations of motion between the two coordinate systems.

*Example* Suppose the material velocity of a particle is  $\mathbf{V} = (Xt, Y - Zt, \sin t)$ , where  $\mathbf{X} = (X, Y, Z)$ . Integrating the equation  $\partial_t \mathbf{U} = \mathbf{V}$ , and using the initial condition  $\mathbf{U}(\mathbf{X}, 0) = \mathbf{0}$ , it follows that

$$\mathbf{U} = \left( \frac{1}{2}Xt^2, Yt - \frac{1}{2}Zt^2, 1 - \cos t \right).$$

To find the spatial versions of the displacement and velocity, recall that  $\mathcal{X} = \mathbf{X} + \mathbf{U}$ . Using our formula for  $\mathbf{U}$ , we have that

$$\mathcal{X} = \left( X\left(1 + \frac{1}{2}t^2\right), Y(1 + t) - \frac{1}{2}Zt^2, Z + 1 - \cos t \right).$$

Taking  $\mathbf{x} = (x, y, z)$ , then, since  $\mathbf{x} = \mathfrak{X}$ , we get that  $x = X(1 + \frac{1}{2}t^2)$ . Solving this yields  $X = 2x/(2 + t^2)$ . In a similar manner, one finds that  $Y = (y + \frac{1}{2}t^2(z - 1 + \cos t))/(1 + t)$  and  $Z = z - 1 + \cos t$ . According to (8.7),  $\mathbf{u}$  is obtained by substituting these formulas into  $\mathbf{U}$ . Doing this, it follows that

$$\mathbf{u} = \left( \frac{xt^2}{2+t^2}, \frac{t}{1+t} \left( y - \frac{1}{2}t(z - 1 + \cos t) \right), 1 - \cos t \right).$$

In a similar manner, using (8.8),

$$\mathbf{v} = \left( \frac{2xt}{2+t^2}, \frac{y}{1+t} - \frac{1}{2}t(z - 1 + \cos t) \frac{2+t}{1+t}, \sin t \right). \quad \blacksquare$$

### 8.2.1 Deformation Gradient

The assumption that the transformation between the two coordinate systems is invertible is one of the fundamental hypotheses in continuum mechanics. To explore this a bit more, suppose that given a material point  $\mathbf{X}_0$  that its spatial counterpart is  $\mathbf{x}_0 = \mathfrak{X}(\mathbf{X}_0, t)$ . For material points  $\mathbf{X} = \mathbf{X}_0 + \Delta\mathbf{X}$  near  $\mathbf{X}_0$ , we have from Taylor's theorem

$$\begin{aligned} \mathbf{x} &= \mathfrak{X}(\mathbf{X}_0 + \Delta\mathbf{X}, t) \\ &\approx \mathfrak{X}(\mathbf{X}_0, t) + \mathbf{F}\Delta\mathbf{X} \\ &= \mathbf{x}_0 + \mathbf{F}\Delta\mathbf{X}, \end{aligned} \tag{8.9}$$

where  $\mathbf{F}$  is the Jacobian matrix for  $\mathfrak{X}$ , evaluated at  $\mathbf{X}_0$ . Letting  $\mathfrak{X} = (\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3)$  and  $\mathbf{X} = (X, Y, Z)$ , then

$$\mathbf{F} = \begin{pmatrix} \frac{\partial \mathfrak{X}_1}{\partial X} & \frac{\partial \mathfrak{X}_1}{\partial Y} & \frac{\partial \mathfrak{X}_1}{\partial Z} \\ \frac{\partial \mathfrak{X}_2}{\partial X} & \frac{\partial \mathfrak{X}_2}{\partial Y} & \frac{\partial \mathfrak{X}_2}{\partial Z} \\ \frac{\partial \mathfrak{X}_3}{\partial X} & \frac{\partial \mathfrak{X}_3}{\partial Y} & \frac{\partial \mathfrak{X}_3}{\partial Z} \end{pmatrix}, \tag{8.10}$$

or, in operator form,  $\mathbf{F} = \nabla_{\mathbf{X}}\mathfrak{X}$ . This matrix plays an important role in continuum mechanics and is known as the *deformation gradient*. The reason it is important can be seen in (8.9), which shows that as a local approximation,  $\mathbf{x} - \mathbf{x}_0 = \mathbf{F}(\mathbf{X} - \mathbf{X}_0)$ . Consequently,  $\mathbf{F}$  is a measure of how much the motion is rotating and distorting the material. We will return to this idea later in the chapter, in Sect. 8.12.2, once the equations of motion are derived.

It is assumed that the transformation between the spatial and material coordinate systems is invertible. More specifically, it is assumed that it is possible to solve  $\mathbf{x} - \mathbf{x}_0 = \mathbf{F}(\mathbf{X} - \mathbf{X}_0)$  for  $\mathbf{X}$ . The result is that  $\mathbf{X} = \mathbf{X}_0 + \mathbf{F}^{-1}(\mathbf{x} - \mathbf{x}_0)$ . The requirement for this to hold is that  $\mathbf{F}$  is invertible, in other words  $\det(\mathbf{F}) \neq 0$ . This means that  $\det(\mathbf{F})$  is either always positive or it is always negative. Given that  $\mathcal{X}(\mathbf{X}, 0) = \mathbf{X}$ , so  $\mathbf{F} = \mathbf{I}$  at  $t = 0$ , then the requirement is that

$$\det(\mathbf{F}) > 0. \quad (8.11)$$

The one-dimensional version of this condition is given in (6.19). As in Chap. 6, this inequality is assumed to hold whenever discussing the continuum theory.

*Example (Uniform Dilatation)* A motion given by

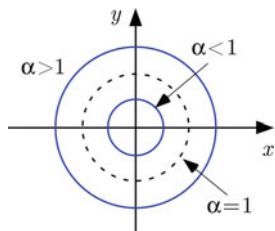
$$\mathbf{x} = \alpha(t)\mathbf{X}, \quad (8.12)$$

where  $\alpha(0) = 1$ , is called a uniform dilatation. To explain how it gets this name, suppose we start out with a sphere of radius  $r$  that is centered at the origin. So, the  $\mathbf{X}$ 's satisfy  $\|\mathbf{X}\| = r$ . According to (8.12), at later times we still have a sphere, centered at the origin, but with a radius  $\alpha r$ . If  $\alpha > 1$  there is a uniform expansion while for  $0 < \alpha < 1$  there is uniform contraction. This is illustrated in Fig. 8.1 for a circular region in the plane.

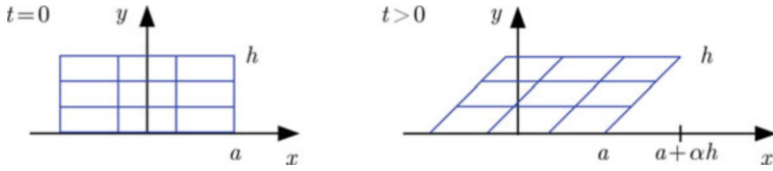
Calculating the displacement and velocity, in material coordinates, using (8.1) and (8.2), we have that  $\mathbf{U}(\mathbf{X}, t) = (\alpha - 1)\mathbf{X}$  and  $\mathbf{V}(\mathbf{X}, t) = \alpha'\mathbf{X}$ . To find the spatial version, we solve (8.12) to obtain  $\mathbf{X} = \mathbf{x}/\alpha$ . From (8.7) and (8.8) it follows that  $\mathbf{u}(\mathbf{x}, t) = (\alpha - 1)\mathbf{x}/\alpha$  and  $\mathbf{v}(\mathbf{x}, t) = \alpha'\mathbf{x}/\alpha$ . Therefore, as in the one-dimensional case,  $\mathbf{v} \neq \frac{\partial \mathbf{u}}{\partial t}$ . Finally, from (8.10), the deformation gradient is  $\mathbf{F} = \alpha\mathbf{I}$ . To satisfy the impenetrability of matter condition (8.11), it is required that  $\alpha > 0$ . ■

*Example (Simple Shear)* A motion given by  $x = X + \alpha(t)Y$ ,  $y = Y$ ,  $z = Z$ , where  $\alpha(0) = 0$ , is an example of simple shear. An illustration of what happens in simple shear is shown in Fig. 8.2, where a rectangle is transformed into a parallelogram with the same height. This is the type of motion one gets when pushing on the side of a deck of cards. To determine the kinematic variables, because  $\mathcal{X} = (X + \alpha(t)Y, Y, Z)$ , then, from (8.1), we have that  $\mathbf{U}(\mathbf{X}, t) = (\alpha(t)Y, 0, 0)$ . To find the displacement in spatial coordinates, we solve  $\mathbf{x} = (X + \alpha(t)Y, Y, Z)$  to

**Fig. 8.1** Uniform dilatation of a circular region







**Fig. 8.2** Simple shear of a rectangular region

obtain  $(X, Y, Z) = (x - \alpha y, y, z)$ . With this,

$$\begin{aligned}\mathbf{u}(\mathbf{x}, t) &= \mathbf{U}(x - \alpha y, y, z, t) \\ &= (\alpha(t)y, 0, 0).\end{aligned}$$

Finally, from (8.10)

$$\mathbf{F} = \begin{pmatrix} 1 & \alpha & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Given that  $\det(\mathbf{F}) = 1$ , then this motion satisfies the impenetrability of matter condition for any value of  $\alpha$ . ■

*Example (Rigid Body Motion)* A rigid body motion is one given by

$$\mathbf{x} = \mathbf{Q}(t)\mathbf{X} + \mathbf{b}(t), \quad (8.13)$$

where  $\mathbf{Q}(t)$  is a rotation matrix with  $\mathbf{Q}(0) = \mathbf{I}$ , and  $\mathbf{b}(0) = \mathbf{0}$ . Therefore, it consists of a rotation, determined by  $\mathbf{Q}$ , followed by a translation given by  $\mathbf{b}$ . To qualify for a rotation, the matrix  $\mathbf{Q}$  must satisfy  $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$  and  $\det(\mathbf{Q}) = 1$ . As an example, consider a merry-go-round motion, where the points in the  $x, y$ -plane rotate around the  $z$ -axis. This happens if  $\mathbf{b}(t) = \mathbf{0}$  and

$$\mathbf{Q}(t) = \begin{pmatrix} \cos(\omega t) & -\sin(\omega t) & 0 \\ \sin(\omega t) & \cos(\omega t) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (8.14)$$

In this case, the points rotate around the  $z$ -axis with an angular velocity  $\omega$ . ■

### 8.3 Material Derivative

To derive the formula for the material derivative, suppose  $F(\mathbf{X}, t)$  is a variable or function in material coordinates and its spatial version is  $f(\mathbf{x}, t)$ . In this case,  $\frac{\partial F}{\partial t}$  is the time rate of change of the variable for the material point that began at  $\mathbf{X}$ . To determine what this is in spatial coordinates note that  $F$  and  $f$  must produce the same value. Therefore, if the material point that started at  $\mathbf{X}$  is currently located at  $\mathbf{x} = \mathcal{X}(\mathbf{X}, t)$ , then it must be that

$$F(\mathbf{X}, t) = f(\mathcal{X}(\mathbf{X}, t), t). \quad (8.15)$$

Letting  $\mathbf{x} = (x, y, z)$  and  $\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3)$ , we have that

$$\begin{aligned} \frac{\partial F}{\partial t} &= \frac{\partial f}{\partial x} \frac{\partial \mathcal{X}_1}{\partial t} + \frac{\partial f}{\partial y} \frac{\partial \mathcal{X}_2}{\partial t} + \frac{\partial f}{\partial z} \frac{\partial \mathcal{X}_3}{\partial t} + \frac{\partial f}{\partial t} \\ &= \left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right) \cdot \frac{\partial \mathbf{U}}{\partial t} + \frac{\partial f}{\partial t} \\ &= \left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right) \cdot \mathbf{v} + \frac{\partial f}{\partial t} \\ &= \nabla f \cdot \mathbf{v} + \frac{\partial f}{\partial t} \\ &= \left( \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right) f, \end{aligned} \quad (8.16)$$

where

$$\nabla = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right). \quad (8.17)$$

This gives us the next result.

**Material Derivative.** *The material derivative, which is defined as*

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla, \quad (8.18)$$

*is the time rate of change of a function following a material point, but expressed in spatial coordinates.*

It is not hard to show that the material derivative satisfies some, but not all, of the elementary properties of differentiation. For example, for constants  $\alpha$ ,  $\beta$  and functions  $f$ ,  $g$ ,

$$\begin{aligned}\frac{D}{Dt}(\alpha f + \beta g) &= \alpha \frac{Df}{Dt} + \beta \frac{Dg}{Dt}, \\ \frac{D}{Dt}(fg) &= g \frac{Df}{Dt} + f \frac{Dg}{Dt}.\end{aligned}$$

However, because of the  $\mathbf{v}$  in the formula for the material derivative, it is generally true that

$$\begin{aligned}\frac{D}{Dt} \frac{\partial}{\partial x} &\neq \frac{\partial}{\partial x} \frac{D}{Dt}, \\ \frac{D}{Dt} \frac{\partial}{\partial t} &\neq \frac{\partial}{\partial t} \frac{D}{Dt}.\end{aligned}$$

In other words, interchanging the order of differentiation requires some care when using the material derivative.

A particularly important example is the material derivative of the displacement function. Recalling that  $\mathbf{V} = \frac{\partial}{\partial t}\mathbf{U}$ , it follows from (8.16) that

$$\mathbf{v} = \frac{D\mathbf{u}}{Dt}. \quad (8.19)$$

This is the vector version of (6.13), and some of the complications that arise from this innocent-looking formula are explored in Exercise 8.8.

*Example (Uniform Dilatation (Cont'd))* For uniform dilatation we found that  $\mathbf{u}(\mathbf{x}, t) = (\alpha - 1)\mathbf{x}/\alpha$  and  $\mathbf{v}(\mathbf{x}, t) = \alpha'\mathbf{x}/\alpha$ . To check on (8.19),

$$\begin{aligned}\frac{D\mathbf{u}}{Dt} &= \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{u} \\ &= \frac{\alpha'}{\alpha^2}\mathbf{x} + \frac{\alpha'}{\alpha}\mathbf{x}\left(1 - \frac{1}{\alpha}\right) \\ &= \frac{\alpha'}{\alpha}\mathbf{x}.\end{aligned}$$

So, as expected, (8.19) holds. ■

The above derivation of the material derivative closely follows what was done for one dimension. In fact, this is true for much of what is done in the chapter. There are some notable exceptions to this statement, and this will become evident when we introduce the stress tensor in Sect. 8.6.1.

## 8.4 Mathematical Tools

The key tool in deriving the equations of motion is the Reynolds Transport Theorem. To state this result, consider a collection of material points that at  $t = 0$  occupy a volume  $R(0)$ , as shown in Fig. 8.3. Due to the motion, at later times these same points occupy the volume  $R(t)$ . The surface of this volume is denoted as  $\partial R(t)$ . For example, if  $R$  is the ball  $\|\mathbf{x}\| \leq 2t + 1$ , then  $\partial R(t)$  is the sphere  $\|\mathbf{x}\| = 2t + 1$ . With this, we have the following result.

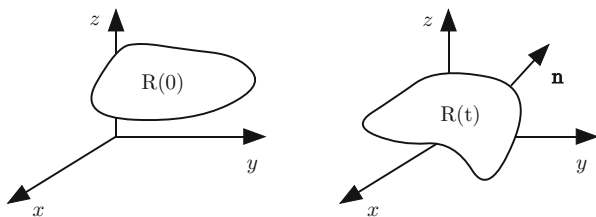
**Reynolds Transport Theorem.** *Assuming  $R(t)$  is regular, and  $f(\mathbf{x}, t)$  is a smooth function, then*

$$\frac{d}{dt} \iiint_{R(t)} f(\mathbf{x}, t) dV = \iiint_{R(t)} \frac{\partial f}{\partial t} dV + \iint_{\partial R(t)} f \mathbf{v} \cdot \mathbf{n} dS, \quad (8.20)$$

where  $\mathbf{n}$  is the unit outward normal to  $R(t)$ ,  $\mathbf{v}$  is the velocity of the points on the surface,  $dV$  is the volume element, and  $dS$  is the surface element.

Stating that  $R(t)$  is regular includes several requirements, all imposed on the original region  $R(0)$  and the motion  $\mathbf{x} = \mathcal{X}(\mathbf{X}, t)$ . First,  $R(0)$  is assumed to satisfy the conditions stated in the integral theorems of multivariable calculus. Namely,  $R(0)$  is bounded with a boundary  $\partial R(0)$  that consists of finitely many smooth, closed orientable surfaces. The second assumption is that the motion is smooth and satisfies (8.11). The reason for this is that in the proof of (8.20) a change of coordinates is made in the volume integral to transform it into an integral over the time-independent domain  $R(0)$ . To use the change of variables theorem from multivariable calculus the Jacobian for the transformation must be nonzero, and that is guaranteed if (8.11) holds.

To outline the proof of (8.20), the first step is to change variables in the integrals so the limits are not dependent on time. The natural choice is to use material coordinates, and let  $\mathbf{x} = \mathcal{X}(\mathbf{X}, t)$ . The Jacobian matrix for this change of variables is  $\mathbf{F}$ , given in (8.10). From the change of variables formula for multiple integrals,



**Fig. 8.3** The material points that occupy the region  $R(0)$  at  $t = 0$  move and at later times occupy  $R(t)$

$$\iiint_{R(t)} f(\mathbf{x}, t) dx dy dz = \iiint_{R(0)} f(\mathbf{X}(\mathbf{X}, t), t) \det(\mathbf{F}) dX dY dZ, \quad (8.21)$$

where  $\det()$  is the determinant.

In the calculations to follow, we need the formula for the derivative of a determinant. This can be derived directly from the definition of a determinant. The result is known as Jacobi's formula, and it states that given a smooth invertible matrix  $\mathbf{M}(t)$ , then

$$\frac{d}{dt} \det(\mathbf{M}) = \det(\mathbf{M}) \operatorname{tr} \left( \mathbf{M}^{-1} \frac{d}{dt} \mathbf{M} \right), \quad (8.22)$$

where  $\operatorname{tr}()$  is the trace. The trace is the sum of the diagonal entries of the matrix, and so  $\operatorname{tr}(\mathbf{M}) = M_{11} + M_{22} + M_{33}$ . Its basic properties, as well as those for the determinant, are given in Appendix D. Setting  $J = \det(\mathbf{F})$ , then from (8.22), and the results from Exercise 8.7,

$$\begin{aligned} \frac{\partial J}{\partial t} &= J \operatorname{tr} \left( \mathbf{F}^{-1} \frac{d}{dt} \mathbf{F} \right) \\ &= J \operatorname{tr} \left( \mathbf{F}^{-1} \nabla_X \mathbf{V} \right) \\ &= J \operatorname{tr} (\nabla \mathbf{v}). \end{aligned} \quad (8.23)$$

In the above expressions, letting  $\mathbf{V} = (V_1, V_2, V_3)$  and  $\mathbf{v} = (v_1, v_2, v_3)$ ,

$$\nabla_X \mathbf{V} = \begin{pmatrix} \frac{\partial V_1}{\partial X} & \frac{\partial V_1}{\partial Y} & \frac{\partial V_1}{\partial Z} \\ \frac{\partial V_2}{\partial X} & \frac{\partial V_2}{\partial Y} & \frac{\partial V_2}{\partial Z} \\ \frac{\partial V_3}{\partial X} & \frac{\partial V_3}{\partial Y} & \frac{\partial V_3}{\partial Z} \end{pmatrix}, \quad (8.24)$$

and

$$\nabla \mathbf{v} = \begin{pmatrix} \frac{\partial v_1}{\partial x} & \frac{\partial v_1}{\partial y} & \frac{\partial v_1}{\partial z} \\ \frac{\partial v_2}{\partial x} & \frac{\partial v_2}{\partial y} & \frac{\partial v_2}{\partial z} \\ \frac{\partial v_3}{\partial x} & \frac{\partial v_3}{\partial y} & \frac{\partial v_3}{\partial z} \end{pmatrix}. \quad (8.25)$$

The above two tensors play an important role in continuum mechanics, and they are called velocity gradients. Specifically,  $\nabla_{\mathbf{X}} \mathbf{V}$  is the *material velocity gradient tensor*, and  $\nabla \mathbf{v}$  is the *spatial velocity gradient tensor*.

One important property that we need here is that

$$\text{tr}(\nabla \mathbf{v}) = \nabla \cdot \mathbf{v}. \quad (8.26)$$

With this, (8.23) reduces to

$$\frac{\partial J}{\partial t} = J(\nabla \cdot \mathbf{v}). \quad (8.27)$$

It should be remembered that in the above expression  $J$  is a function of  $\mathbf{X}$  and  $t$ , and  $\nabla \cdot \mathbf{v}$  is evaluated at  $\mathbf{x} = \mathcal{X}(\mathbf{X}, t)$ .

We are now in a position to differentiate (8.21). Letting  $\mathcal{X} = (X, Y, Z)$ ,

$$\begin{aligned} & \frac{d}{dt} \iiint_{R(t)} f(\mathbf{x}, t) dx dy dz \\ &= \iiint_{R(0)} \left[ \frac{\partial f}{\partial t} J + \frac{\partial f}{\partial x} \frac{\partial X}{\partial t} J + \frac{\partial f}{\partial y} \frac{\partial Y}{\partial t} J + \frac{\partial f}{\partial z} \frac{\partial Z}{\partial t} J + f \frac{\partial J}{\partial t} \right] dX dY dZ \\ &= \iiint_{R(0)} \left[ \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x} V_1 + \frac{\partial f}{\partial y} V_2 + \frac{\partial f}{\partial z} V_3 + f \nabla \cdot \mathbf{v} \right] J dX dY dZ \\ &= \iiint_{R(0)} \left[ \frac{\partial f}{\partial t} + \nabla \mathbf{f} \cdot \mathbf{v} + f \nabla \cdot \mathbf{v} \right] J dX dY dZ \\ &= \iiint_{R(t)} \left[ \frac{\partial f}{\partial t} + \nabla \cdot (f \mathbf{v}) \right] dx dy dz. \end{aligned} \quad (8.28)$$

The next step requires the Divergence Theorem, which states that for a smooth function  $\mathbf{w}$ ,

$$\iiint_R \nabla \cdot \mathbf{w} dV = \iint_{\partial R} \mathbf{w} \cdot \mathbf{n} dS.$$

Taking  $\mathbf{w} = f \mathbf{v}$ , then (8.28) reduces to (8.20), and the theorem is proved.

A useful form of the Reynolds Transport formula comes out of the proof, and it is worth restating the result. From (8.28), and the definition of the material derivative in (8.18), it follows that

$$\frac{d}{dt} \iiint_{R(t)} f(\mathbf{x}, t) dV = \iiint_{R(t)} \left( \frac{Df}{Dt} + f \nabla \cdot \mathbf{v} \right) dV. \quad (8.29)$$

### 8.4.1 General Balance Law

The above integral theorems will be used to take balance laws that are formulated as integrals and express them as differential equations. The steps involved in this derivation are always the same, so it is worth deriving a general formula that can be used when needed. With this in mind, suppose a material set of points occupies a spatial domain  $B(t)$ . Also, suppose that given any volume  $R(t)$  of material points in  $B(t)$ , the following general balance law holds:

$$\frac{d}{dt} \iiint_{R(t)} f(\mathbf{x}, t) dV = - \iint_{\partial R(t)} \mathbf{J} \cdot \mathbf{n} dS + \iiint_{R(t)} Q(\mathbf{x}, t) dV. \quad (8.30)$$

To state the above equation in physical terms,  $f$  can be thought of a density of a quantity, and examples are mass density, momentum density, and energy density. The above balance law states that the rate of change of the total amount of this quantity in a region  $R(t)$  is due to the flux across the boundary and the creation or loss through the volume. The flux in this case is  $\mathbf{J}$ , and  $Q$  is the creation or loss density.

The integral balance law (8.30) is the three-dimensional version of (4.52). From (8.29), and using the Divergence Theorem to convert the surface integral into a volume integral, (8.30) can be written as

$$\iiint_{R(t)} \left( \frac{Df}{Dt} + f \nabla \cdot \mathbf{v} \right) dV = \iiint_{R(t)} [-\nabla \cdot \mathbf{J} + Q(\mathbf{x}, t)] dV.$$

The balance law is assumed to hold for all material volumes  $R(t)$  from  $B(t)$ , and therefore from the du Bois-Reymond Lemma we have that, in  $B(t)$ ,

$$\frac{Df}{Dt} + f \nabla \cdot \mathbf{v} = -\nabla \cdot \mathbf{J} + Q, \quad (8.31)$$

or equivalently

$$\frac{\partial f}{\partial t} + \nabla \cdot (\mathbf{v}f) = -\nabla \cdot \mathbf{J} + Q. \quad (8.32)$$

This equation, in one form or another, has been used repeatedly in this textbook. The one-dimensional version (4.53) was used to derive the diffusion equation in Chap. 4, it was used in the derivation of the traffic flow equation in Chap. 5, and it was used multiple times in Chap. 6. We are now going to use it to derive the equations of continuum mechanics.

### 8.4.2 Direct Notation and Tensors

Using vector operators to express the balance law in (8.32) is an example of what is called direct notation. It can also be written as

$$\frac{\partial f}{\partial t} + \frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y} + \frac{\partial v_3}{\partial z} = -\frac{\partial J_1}{\partial x} - \frac{\partial J_2}{\partial y} - \frac{\partial J_3}{\partial z} + Q,$$

which is the component form of the equation. While the vector form has the advantage of simplicity, the component version is what is usually needed when you actually solve a problem. However, the more important reason for using the vector version is that it holds for any orthogonal coordinate system. We have been using Cartesian coordinates, but it is not unusual to have problems where cylindrical or spherical coordinates are the preferred system. In fact, in the next chapter we will use cylindrical coordinates in multiple examples. What is needed in such cases are the formulas that express the vector operations in the respective coordinate system. The formulas for cylindrical coordinates are given in Appendix E. If you need a more extensive list, you might consult Moon and Spencer (1988).

Another comment worth making has to do with why  $\nabla_X \mathbf{V}$  and  $\nabla \mathbf{v}$  are identified as tensors. A more accurate statement is that they are second-order tensors. In the case of  $\nabla \mathbf{v}$ , this means that the operator represented as  $\nabla \mathbf{v}$  is a linear transformation. The formula given in (8.25) is its matrix representation in the Cartesian coordinate system we are using. If you use a different orthogonal coordinate system, the matrix representation changes accordingly. The fact is that although tensor analysis has played a role in the development of Newtonian mechanics, it is not needed here. What is needed, and what overlaps with introductory tensor analysis, is vector calculus and matrix algebra. Those interested in exploring tensor analysis should consult Aris (1990).

## 8.5 Continuity Equation

The assumption is that mass is neither created nor destroyed. To express this mathematically, assume that at  $t = 0$  a collection of material points occupies the volume  $R(0)$ . At any later time these same points occupy a spatial volume  $R(t)$ . Our assumption means that the total mass of the material in this region does not



change. If we let  $\rho(\mathbf{x}, t)$  designate the mass density of the material (i.e., mass per unit volume), then our assumption states that

$$\frac{d}{dt} \iiint_{R(t)} \rho(\mathbf{x}, t) dV = 0. \quad (8.33)$$

In terms of the general law (8.30), we have that  $f = \rho$ ,  $\mathbf{J} = \mathbf{0}$ , and  $Q = 0$ . Therefore, from (8.31) we have that the continuity equation is

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0. \quad (8.34)$$

### 8.5.1 Incompressibility

When studying the motions of liquids, such as water, it is very often assumed the liquid is incompressible. The idea is that even though a volume of material points moves, and changes shape, the total volume is constant. This assumption provides an addition balance law, and it is that

$$\frac{d}{dt} \iiint_{R(t)} dV = 0. \quad (8.35)$$

In this case, in (8.30),  $f = 1$ ,  $\mathbf{J} = \mathbf{0}$ , and  $Q = 0$ , and so from (8.31) the resulting differential equation is

$$\nabla \cdot \mathbf{v} = 0. \quad (8.36)$$

This is the continuity equation for an incompressible material, fluid or solid. You might be wondering what happens to the more general version given in (8.34). Well, in this case it reduces to  $\frac{D\rho}{Dt} = 0$ . This states that the density following a material point does not change in time. Therefore, if the density is initially constant, then it is constant for all time. In this textbook, whenever discussing an incompressible material it will always be assumed that the initial density is constant, so  $\rho$  is constant.

*Example (Uniform Dilatation)* For uniform dilatation,  $\mathbf{v}(\mathbf{x}, t) = \alpha' \mathbf{x} / \alpha$ . In this case,  $\nabla \cdot \mathbf{v} = 3\alpha' / \alpha$ . Assuming that  $\alpha$  is not constant, then  $\nabla \cdot \mathbf{v} \neq 0$ . Therefore, uniform dilatation is not possible for an incompressible material. ■

*Example (Translational Motion)* For translational motion the velocity  $\mathbf{v}$  is independent of  $\mathbf{x}$ , but can depend on  $t$ . Given that  $\mathbf{v} = \mathbf{v}(t)$ , then  $\nabla \cdot \mathbf{v} = 0$ . This means that translational motion is possible for an incompressible material. This conclusion makes sense from a physical point-of-view because the volume of an

object is unaffected by translation. By the same reasoning, it is expected that rigid body motion is possible for an incompressible material, and this is proved in Exercise 8.5. ■

The assumption of incompressibility is often made, and it has the benefit of usually making the problem easier to solve. However, it is not at all obvious what conditions must hold for this assumption to be appropriate. This issue will be addressed in more detail once the momentum equations have been derived.

## 8.6 Linear Momentum Equation

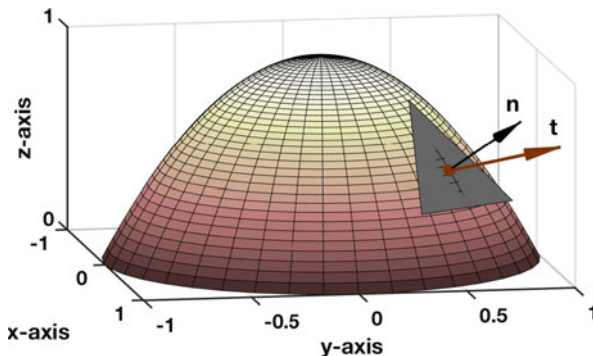
The derivation of the momentum equation requires more effort than for mass conservation. One complication is that we now need to distinguish between linear and angular momentum. We start with linear momentum, which is what was derived in the one-dimensional formulation in Chap. 6.

As in Sect. 8.4.1, the material is assumed to occupy a spatial domain  $B(t)$ , and  $R(t)$  is a volume of material points within  $B(t)$ . It is assumed that the material in  $R(t)$  is subject to external body forces, measured with respect to unit mass, and these are denoted as  $\mathbf{f}$ . There are also forces that act on the surface of  $R$  due to the relative deformation of the material. The one-dimensional version of this is shown in Fig. 6.3. The idea is that the material points in  $B$ , that are external to  $R$ , act on  $R$  across the surface  $\partial R$ . To incorporate this into the balance law, given a point  $\mathbf{x}$  on the surface  $\partial R$ , let  $\mathbf{t}$  be the force, per unit area, on  $R$  due to the material exterior to  $R$ . Because of its units,  $\mathbf{t}$  is referred to as a stress vector. This is illustrated in Fig. 8.4.

With this, the balance of linear momentum gives us that

$$\frac{d}{dt} \iiint_{R(t)} \rho \mathbf{v} dV = \iint_{\partial R(t)} \mathbf{t} dS + \iiint_{R(t)} \rho \mathbf{f} dV. \quad (8.37)$$

**Fig. 8.4** A triangular piece of the plane tangent to the surface, along with the unit outward normal  $\mathbf{n}$  and stress vector  $\mathbf{t}$



It is important to understand that the above equation is an assumption, and is one of the balance laws of continuum mechanics. We will reduce it to a differential equation, but before doing that it is necessary to consider the stress vector  $\mathbf{t}$  in more detail.

### 8.6.1 Stress Tensor

A complicating factor in (8.37) is that the stress  $\mathbf{t}$  depends on direction. As an example, if you pull on a sheet of rubber, the stress in the direction of the pull is different than the stress perpendicular to the pull. To explain the implications of this, given that we are dealing with vectors in  $\mathbb{R}^3$ , at any given point  $\bar{\mathbf{x}}$  on the surface  $\partial R$  in (8.37), we should be able to define, or be able to determine, three basis vectors  $\boldsymbol{\sigma}_x, \boldsymbol{\sigma}_y, \boldsymbol{\sigma}_z$  so that  $\mathbf{t} = a\boldsymbol{\sigma}_x + b\boldsymbol{\sigma}_y + c\boldsymbol{\sigma}_z$ . Moreover, given any other surface that contains the same point  $\bar{\mathbf{x}}$ , we should be able to use the same basis vectors and write  $\mathbf{t} = \bar{a}\boldsymbol{\sigma}_x + \bar{b}\boldsymbol{\sigma}_y + \bar{c}\boldsymbol{\sigma}_z$ . Because of the dependence on direction, without some other piece of information, there will be no connection, in general, between the coefficients  $a, b, c$  and  $\bar{a}, \bar{b}, \bar{c}$ . Fortunately, for us,  $\mathbf{t}$  must satisfy the balance of linear momentum equation (8.37). The resulting connection, which is given below, is often stated to be the most important theorem in continuum mechanics (e.g., see Gurtin and Martins 1976).

To set the stage, note that applying the Reynolds Transport Theorem, as expressed in (8.29), the balance law in (8.37) can be written in the form

$$\iint_{\partial R} \mathbf{t} dS + \iiint_R \mathbf{b} dV = \mathbf{0},$$

where  $\mathbf{b}$  includes all of the terms involving the volume integral. The key observation is that this must hold for any regular region  $R$  contained in the spatial domain  $B(t)$  occupied by the material. Specifically, it holds for every tetrahedron  $D$ , and to make this explicit, we write

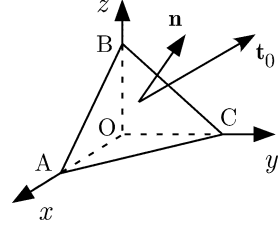
$$\iint_{\partial D} \mathbf{t} dS + \iiint_D \mathbf{b} dV = \mathbf{0}. \quad (8.38)$$

This brings us to the next result.

**Cauchy Stress Theorem.** *Assume that  $\mathbf{t}$  and  $\mathbf{b}$  are continuous, and (8.38) holds for every tetrahedron  $D$  in  $B$ . In this case, there exists a tensor  $\boldsymbol{\sigma}$ , known as the Cauchy stress tensor, with the property that given any smooth surface  $\partial R$  in  $B$ , with unit normal  $\mathbf{n}$ ,  $\mathbf{t} = \boldsymbol{\sigma} \mathbf{n}$ .*

To outline the proof, given a point  $\mathbf{x}$  in  $B$ , we will consider tetrahedra positioned so  $\mathbf{x}$  is in one of the faces, and the other faces are aligned with the coordinate axes. This is illustrated in Fig. 8.5, and this is referred to as a Cauchy tetrahedron. As a

**Fig. 8.5** Cauchy tetrahedron used to derive the stress tensor in (8.41)



point on this triangular surface, there is an outward normal  $\mathbf{n}$ , and the corresponding stress vector  $\mathbf{t}_0$ . Because the stress is assumed to be continuous, and assuming the face  $ABC$  is relatively small,

$$\iint_{ABC} \mathbf{t} dS \approx \mathbf{t}_0 \Delta A,$$

where  $\Delta A$  is the area of the face.

Similar approximations hold for the other three faces. For example, given a point in  $AOB$ , the tetrahedron (so,  $y > 0$ ) exerts a stress  $\boldsymbol{\sigma}_y = (\sigma_{12}, \sigma_{22}, \sigma_{32})^T$  on the material on the other side of the face (where  $y < 0$ ). Using Newton's third law, it follows that the stress on the tetrahedron, at the given point on  $AOB$ , is  $-\boldsymbol{\sigma}_y$ . With this, we have the approximation

$$\iint_{AOB} \mathbf{t} dS \approx -\boldsymbol{\sigma}_y \Delta A_y,$$

where  $\Delta A_y$  is the area of the face. There are similar stress vectors,  $\boldsymbol{\sigma}_x = (\sigma_{11}, \sigma_{21}, \sigma_{31})^T$  and  $\boldsymbol{\sigma}_z = (\sigma_{13}, \sigma_{23}, \sigma_{33})^T$ , for the other two faces.

Using a similar approximation for the volume integral, we have that a first-order approximation for (8.38) is

$$-\Delta A_x \boldsymbol{\sigma}_x - \Delta A_y \boldsymbol{\sigma}_y - \Delta A_z \boldsymbol{\sigma}_z + \Delta A \mathbf{t}_0 + \mathbf{b} \Delta V = \mathbf{0}, \quad (8.39)$$

where  $\mathbf{b}$  is evaluated at a point within the tetrahedron, and  $\Delta V$  is its volume. Now, using the properties of a cross-product, one can show that the areas of the four faces of the tetrahedron are related as follows:  $\Delta A_x = n_1 \Delta A$ ,  $\Delta A_y = n_2 \Delta A$ ,  $\Delta A_z = n_3 \Delta A$ . Also,  $\Delta V = \frac{1}{3} h \Delta A$ , where  $h$  is the height of the tetrahedron. Introducing these into (8.39), dividing by  $\Delta A$ , and letting the tetrahedron shrink to zero (so,  $h$  goes to zero), we have that

$$\begin{aligned} \mathbf{t}_0 &= n_1 \boldsymbol{\sigma}_x + n_2 \boldsymbol{\sigma}_y + n_3 \boldsymbol{\sigma}_z \\ &= n_1 \begin{pmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{31} \end{pmatrix} + n_2 \begin{pmatrix} \sigma_{12} \\ \sigma_{22} \\ \sigma_{32} \end{pmatrix} + n_3 \begin{pmatrix} \sigma_{13} \\ \sigma_{23} \\ \sigma_{33} \end{pmatrix} \\ &= \boldsymbol{\sigma} \mathbf{n}, \end{aligned} \quad (8.40)$$

where

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \quad (8.41)$$

is known as the *Cauchy stress tensor*. This proves that three stress vectors can be used to determine the stress in any direction. One consequence of this is that we have nine unknown stress functions in (8.41). As in Chap. 6, we will use a constitutive law to determine these functions.

Not everything was explained in the above derivation, and this was done to simplify the presentation. The approximations used for the integrals are effectively first term Taylor series approximations. However, given that the stress is only assumed to be continuous, it is more appropriate to invoke the mean value theorem for integrals. Those interested in a more formal proof, or an extension of this result to include more general situations, such as might arise if there were shock solutions, should consult Gurtin et al. (1968) and Gurtin and Martins (1976).

### 8.6.2 Differential Form of Equation

Using (8.40), the balance law (8.37) for linear momentum becomes

$$\frac{d}{dt} \iiint_{R(t)} \rho \mathbf{v} dV = \iint_{\partial R(t)} \boldsymbol{\sigma} \mathbf{n} dS + \iiint_{R(t)} \rho \mathbf{f} dV. \quad (8.42)$$

This is now in a form that is the same as the general law in (8.30). Using (8.31) we therefore conclude that

$$\frac{D}{Dt}(\rho \mathbf{v}) + (\nabla \cdot \mathbf{v})(\rho \mathbf{v}) = \nabla \cdot \boldsymbol{\sigma} + \rho \mathbf{f}. \quad (8.43)$$

The divergence of  $\boldsymbol{\sigma}$  in the above equation is defined as

$$\nabla \cdot \boldsymbol{\sigma} = \begin{pmatrix} \frac{\partial \sigma_{11}}{\partial x} + \frac{\partial \sigma_{12}}{\partial y} + \frac{\partial \sigma_{13}}{\partial z} \\ \frac{\partial \sigma_{21}}{\partial x} + \frac{\partial \sigma_{22}}{\partial y} + \frac{\partial \sigma_{23}}{\partial z} \\ \frac{\partial \sigma_{31}}{\partial x} + \frac{\partial \sigma_{32}}{\partial y} + \frac{\partial \sigma_{33}}{\partial z} \end{pmatrix}. \quad (8.44)$$

Expanding the material derivative, and using the continuity equation (8.34), we obtain

$$\rho \frac{D\mathbf{v}}{Dt} = \nabla \cdot \boldsymbol{\sigma} + \rho \mathbf{f}. \quad (8.45)$$

This is the equation for linear momentum, or just the momentum equation for short. It is expressed in spatial coordinates, and the material coordinates version is given in Sect. 8.12.

## 8.7 Angular Momentum

Unlike continuity and linear momentum, the equation for angular momentum is simply the statement that the stress tensor is symmetric. To obtain this result, we consider the angular momentum of the volume  $R(t)$ . For a single point the angular momentum per unit volume is  $\mathbf{x} \times (\rho \mathbf{v})$ . Integrating this over the volume, and accounting for the same forces used for linear momentum, it follows that the balance law for angular momentum is

$$\frac{d}{dt} \iiint_{R(t)} \mathbf{x} \times (\rho \mathbf{v}) dV = \iint_{\partial R(t)} \mathbf{x} \times (\boldsymbol{\sigma} \mathbf{n}) dS + \iiint_{R(t)} \mathbf{x} \times (\rho \mathbf{f}) dV. \quad (8.46)$$

Carrying out the cross products, writing the result in our standard balance law format, and then using the linear momentum equation to simplify the expression, the resulting equation is  $(\sigma_{32} - \sigma_{23}, \sigma_{13} - \sigma_{31}, \sigma_{21} - \sigma_{12})^T = \mathbf{0}$ . The conclusion is that  $\sigma_{32} = \sigma_{23}$ ,  $\sigma_{13} = \sigma_{31}$ , and  $\sigma_{21} = \sigma_{12}$ . Therefore, as stated earlier, to satisfy the balance law for angular momentum,  $\boldsymbol{\sigma}$  must be symmetric.

## 8.8 Summary of the Equations of Motion

To summarize the equations of motion up to this point, we have found that the continuity and momentum equations are, respectively,

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0, \quad (8.47)$$

$$\rho \frac{D\mathbf{v}}{Dt} = \nabla \cdot \boldsymbol{\sigma} + \rho \mathbf{f}, \quad (8.48)$$

where  $\boldsymbol{\sigma}$  is symmetric. If the material is assumed to incompressible, and the initial density is constant, then (8.47) is replaced with

$$\nabla \cdot \mathbf{v} = 0, \quad (8.49)$$

and  $\rho$  is a constant. Depending on the region the material occupies, boundary and initial conditions must be supplied to complete the problem.

Although the above equations are quite general, certain assumptions were made in the derivation. In particular, it was assumed that mass is not created or destroyed. If this occurs, then both the continuity and the momentum equations are affected. We also assumed that there are no sources of angular momentum, other than what comes from the linear momentum forcing function  $\mathbf{f}$ . There are situations where this does not happen, and the most well known are micropolar materials. Those interested in investigating what this means in terms of the model and analysis should consult Eringen (2001).

One last comment worth making here is that the above equations are coordinate free. This means that if a particular orthogonal coordinate system is preferred, such as cylindrical coordinates, one only needs to determine the formulas for the divergence and gradient operators to be able to determine the equations of motion. An example of this will be given in Sect. 9.1.2.

### 8.8.1 The Assumption of Incompressibility

The idea of something being incompressible is easy to understand. However, in the real world everything can be reduced in volume, you just need to use a large enough stress. So, incompressibility is an approximation. What is not obvious is when this approximation can be used. One approach is to consider how much the volume changes as you increase the pressure on a material. To quantify this statement, suppose a material subjected to a pressure  $p_0$  occupies a volume  $v_0$ , and at a larger pressure  $p_1$  it occupies a (smaller) volume  $v_1$ . This is used to define the *bulk modulus*  $K$ , which is

$$K = -\frac{p_1 - p_0}{(v_1 - v_0)/v_0}.$$

This means that for an incompressible material, so  $v_1 = v_0$ ,  $K = \infty$ . It is found experimentally that for air  $K = 10^{-4}$  GPa, for water  $K = 2$  GPa, and for stainless steel  $K = 160$  GPa. Whether you can consider these incompressible depends on how  $K$  compares with typical pressure, or stress, differences in the problem. For example, suppose you fill a gallon paint can with water and then stand on the lid (which is assumed to be moveable). Since  $K = 2$  GPa, you would have to weigh about 96,000 lbs (44,000 kg) to compress the water volume just 1%. So, for most everyday applications, water can be considered to be incompressible. To check on air, the pressure created by a steady wind is  $p = \frac{1}{2}\rho v^2$  (this formula is derived in Sect. 9.3.1). Based on this, a wind speed of about 200 mph (330 km/h) is necessary to compress the volume of air by 5%. Consequently, unless you are considering problems with fairly high wind speeds, the assumption of incompressibility is probably appropriate.

A related question is whether it is possible to have a material that is slightly compressible. The idea is that incompressibility provides a first term approximation, and then you find a second term in the expansion of the solution that accounts for the effects of compressibility. This is known as the “incompressible limit” question, and it has been studied extensively for fluids, and to a lesser extend for solids. Those interested in learning about this should consult Alazard (2006), Schochet (1985), or Destrade et al. (2002).

## 8.9 Constitutive Laws

It remains to specify the constitutive law for the stress. As discussed in Chap. 6, there are certain requirements these laws are expected to satisfy. We are interested here in only one, and it is the Principle of Material Frame-Indifference. The basic requirement is that the stress does not depend on the observer, assuming observers are connected by a rigid body motion.

To put this into a mathematical framework, suppose we change coordinates from  $\mathbf{x}$  to  $\mathbf{x}^*$  using rigid body motion. Specifically,

$$\mathbf{x}^* = \mathbf{Q}(t)\mathbf{x} + \mathbf{b}(t), \quad (8.50)$$

where  $\mathbf{Q}(t)$  is a rotation matrix, and  $\mathbf{Q}$  and  $\mathbf{b}$  are smooth functions of time. To qualify for a rotation, the matrix  $\mathbf{Q}$  must satisfy  $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$  and  $\det(\mathbf{Q}) = 1$ . An example of such a matrix is given in (8.14). One important property of rotations is that  $\mathbf{Q}^{-1} = \mathbf{Q}^T$ , and this is a direct consequence of the equation  $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ . In the parlance of continuum mechanics, (8.50) is known as an Euclidean transformation. It differs from a Galilean transformation, often studied in Newtonian physics. This is because for a Galilean transformation  $\mathbf{Q}$  is taken to be constant and  $\mathbf{b}$  is linear in time. Finally, as used in Sect. 6.10.1, the general version of a Euclidean transformation includes a time translation  $t^* = t - t_0$ . Given the role of the reference frame in defining the stress, for simplicity, it is assumed in what follows that  $t_0 = 0$ .

The rigid body motion assumption has consequences for the material coordinate system. Given that the spatial system reduces to the material system when  $t = 0$ , then from (8.50) we have that

$$\mathbf{X}^* = \mathbf{Q}_0\mathbf{X} + \mathbf{b}_0, \quad (8.51)$$

where  $\mathbf{Q}_0 = \mathbf{Q}(0)$  is a rotation, and both  $\mathbf{Q}_0$  and  $\mathbf{b}_0 = \mathbf{b}(0)$  are constants. This means that if our two observers want to revert to material coordinates, the above expression tells them how their material coordinate systems are related.

There are two tenets of the Principle of Material Frame-Indifference, and both are assumptions on the properties of the stress when measured by different observers.



*Tenet 1: Objectivity*

Given basis vectors  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$  in the  $\mathbf{x}$  system, the corresponding basis vectors in the  $\mathbf{x}^*$  system are  $\mathbf{e}_1^* = \mathbf{Q}\mathbf{e}_1, \mathbf{e}_2^* = \mathbf{Q}\mathbf{e}_2, \mathbf{e}_3^* = \mathbf{Q}\mathbf{e}_3$ . Using the respective basis vectors, the stress has a representation  $\boldsymbol{\sigma}$  in the  $\mathbf{x}$  system, and it has a representation  $\boldsymbol{\sigma}^*$  in the  $\mathbf{x}^*$  system. The Principle of Frame-Indifference requires that the stress obeys the usual change of basis formula from linear algebra, and so the requirement is that

$$\boldsymbol{\sigma}^* = \mathbf{Q}\boldsymbol{\sigma}\mathbf{Q}^T. \quad (8.52)$$

A tensor  $\boldsymbol{\sigma}$  that satisfies this equation is said to be *objective*, or Euclidean frame-indifferent. Put into words, (8.52) states that the stress in the  $\mathbf{x}^*$  system can be found by rotating back to the  $\mathbf{x}$  system, calculating the stress there, and then rotating the result over to the  $\mathbf{x}^*$  system.

*Tenet 2: Form Invariance*

The second tenet of Material Frame-Indifference concerns the functional form of the constitutive law. To explain, suppose that the proposed constitutive law for the stress states that it depends on a quantity  $\mathbf{R}$ . In other words, there is a function  $\mathbf{G}$  so that

$$\boldsymbol{\sigma} = \mathbf{G}(\mathbf{R}). \quad (8.53)$$

The assumption is that the form of the constitutive law does not depend on the observer. Therefore, if  $\boldsymbol{\sigma}^*$  and  $\mathbf{R}^*$  are the  $\mathbf{x}^*$  system versions of  $\boldsymbol{\sigma}$  and  $\mathbf{R}$ , then the requirement is that

$$\boldsymbol{\sigma}^* = \mathbf{G}(\mathbf{R}^*). \quad (8.54)$$

In this case the constitutive law is said to be *form invariant*.

To summarize the above discussion, the requirement on the constitutive law coming from the Principle of Frame-Indifference is that  $\boldsymbol{\sigma}^*$  coming from (8.54) must satisfy (8.52) for all  $\mathbf{Q}$  and  $\mathbf{b}$ , as used in (8.50). Moreover, the constitutive law must be consistent with the requirement that  $\boldsymbol{\sigma}$  is symmetric.

To use the above requirements, we need some basic information about how the variables transform under a rigid body motion. In the  $\mathbf{x}$  coordinate system, the material points move according to the rule  $\mathbf{x} = \mathcal{X}(\mathbf{X}, t)$ , while in the  $\mathbf{x}^*$  coordinate system the material points move according to the rule  $\mathbf{x}^* = \mathcal{X}^*(\mathbf{X}^*, t)$ . Given (8.50), it follows that  $\mathcal{X}^* = \mathbf{Q}\mathcal{X} + \mathbf{b}$ . Taking the time derivative of this equation it follows that

$$\mathbf{v}^* = \mathbf{Q}\mathbf{v} + \mathbf{Q}'\mathcal{X} + \mathbf{b}',$$

or equivalently

$$\mathbf{v}^* = \mathbf{Q}\mathbf{v} + \mathbf{Q}'\mathbf{x} + \mathbf{b}'. \quad (8.55)$$

In a similar manner it follows that the deformation gradient, given in (8.10), transforms as

$$\mathbf{F}^* = \mathbf{Q}\mathbf{F}\mathbf{Q}_0^T. \quad (8.56)$$

This shows that  $\mathbf{F}$  does not satisfy (8.52), and it is therefore not objective.

*Example 1* Suppose that it is assumed  $\boldsymbol{\sigma} = \lambda \mathbf{F}^T \mathbf{F}$ . First, note that the symmetry condition is satisfied. As for frame-indifference, according to (8.54),  $\boldsymbol{\sigma}^* = \lambda (\mathbf{F}^*)^T \mathbf{F}^*$ . Using (8.56),

$$\begin{aligned} (\mathbf{F}^*)^T \mathbf{F}^* &= (\mathbf{Q}\mathbf{F}\mathbf{Q}_0^T)^T (\mathbf{Q}\mathbf{F}\mathbf{Q}_0^T) \\ &= (\mathbf{Q}_0 \mathbf{F}^T \mathbf{Q}^T) (\mathbf{Q}\mathbf{F}\mathbf{Q}_0^T) \\ &= \mathbf{Q}_0 \mathbf{F}^T \mathbf{F} \mathbf{Q}_0^T. \end{aligned}$$

So,  $\boldsymbol{\sigma}^* = \lambda \mathbf{Q}_0 \mathbf{F}^T \mathbf{F} \mathbf{Q}_0^T$ . To satisfy (8.52), we need

$$\mathbf{Q}_0 \mathbf{F}^T \mathbf{F} \mathbf{Q}_0^T = \mathbf{Q} \mathbf{F}^T \mathbf{F} \mathbf{Q}^T.$$

It is not always true that  $\mathbf{Q}(t) = \mathbf{Q}(0)$ , and an example is given in (8.14). This means that (8.52) is not satisfied for all rotation matrices  $\mathbf{Q}(t)$ . Therefore, this constitutive law does not satisfy the Principle of Frame-Indifference. ■

*Example 2* The equation for the velocity in (8.55) is an example of a function in the  $\mathbf{x}^*$  system equal to one in the  $\mathbf{x}$  system. The general version of this is  $\mathbf{g}^*(\mathbf{x}^*, t) = \mathbf{h}(\mathbf{x}, t)$ . To relate the derivatives of these two functions, from (8.50) we have that

$$\mathbf{x} = \mathbf{Q}^T (\mathbf{x}^* - \mathbf{b}).$$

Letting  $\mathbf{g}^* = (g_1^*, g_2^*, g_3^*)$  and  $\mathbf{h} = (h_1, h_2, h_3)$ , then using the chain rule

$$\begin{aligned} \frac{\partial g_1^*}{\partial x^*} &= \frac{\partial h_1}{\partial x} \frac{\partial x}{\partial x^*} + \frac{\partial h_1}{\partial y} \frac{\partial y}{\partial x^*} + \frac{\partial h_1}{\partial z} \frac{\partial z}{\partial x^*} \\ &= \frac{\partial h_1}{\partial x} Q_{11} + \frac{\partial h_1}{\partial y} Q_{12} + \frac{\partial h_1}{\partial z} Q_{13}. \end{aligned}$$

Carrying out the other derivatives the conclusion is that

$$(\nabla \mathbf{g})^* = (\nabla \mathbf{h}) \mathbf{Q}^T,$$

where

$$(\nabla \mathbf{g})^* = \begin{pmatrix} \frac{\partial g_1^*}{\partial x^*} & \frac{\partial g_1^*}{\partial y^*} & \frac{\partial g_1^*}{\partial z^*} \\ \frac{\partial g_2^*}{\partial x^*} & \frac{\partial g_2^*}{\partial y^*} & \frac{\partial g_2^*}{\partial z^*} \\ \frac{\partial g_3^*}{\partial x^*} & \frac{\partial g_3^*}{\partial y^*} & \frac{\partial g_3^*}{\partial z^*} \end{pmatrix}, \quad \nabla \mathbf{h} = \begin{pmatrix} \frac{\partial h_1}{\partial x} & \frac{\partial h_1}{\partial y} & \frac{\partial h_1}{\partial z} \\ \frac{\partial h_2}{\partial x} & \frac{\partial h_2}{\partial y} & \frac{\partial h_2}{\partial z} \\ \frac{\partial h_3}{\partial x} & \frac{\partial h_3}{\partial y} & \frac{\partial h_3}{\partial z} \end{pmatrix}.$$

As an example, with the velocity given (8.55),

$$\begin{aligned} (\nabla \mathbf{v})^* &= (\mathbf{Q} \nabla \mathbf{v} + \mathbf{Q}') \mathbf{Q}^T \\ &= \mathbf{Q} (\nabla \mathbf{v}) \mathbf{Q}^T + \mathbf{Q}' \mathbf{Q}^T. \quad \blacksquare \end{aligned} \quad (8.57)$$

*Example 3* Suppose it is assumed that  $\boldsymbol{\sigma} = \mu [\nabla \mathbf{v} + (\nabla \mathbf{v})^T]$ , where  $\mu$  is a constant. In this case,  $\boldsymbol{\sigma}^* = \mu [(\nabla \mathbf{v})^* + ((\nabla \mathbf{v})^*)^T]$ . Using (8.57),

$$\begin{aligned} (\nabla \mathbf{v})^* + ((\nabla \mathbf{v})^*)^T &= \mathbf{Q} (\nabla \mathbf{v}) \mathbf{Q}^T + \mathbf{Q}' \mathbf{Q}^T + [\mathbf{Q} (\nabla \mathbf{v}) \mathbf{Q}^T + \mathbf{Q}' \mathbf{Q}^T]^T \\ &= \mathbf{Q} (\nabla \mathbf{v}) \mathbf{Q}^T + \mathbf{Q}' \mathbf{Q}^T + \mathbf{Q} (\nabla \mathbf{v})^T \mathbf{Q}^T + \mathbf{Q} (\mathbf{Q}')^T \\ &= \mathbf{Q} [(\nabla \mathbf{v}) + (\nabla \mathbf{v})^T] \mathbf{Q}^T + \mathbf{Q}' \mathbf{Q}^T + \mathbf{Q} (\mathbf{Q}')^T. \end{aligned}$$

Since  $\mathbf{Q}' \mathbf{Q}^T + \mathbf{Q} (\mathbf{Q}')^T = \frac{d}{dt} (\mathbf{Q} \mathbf{Q}^T) = \frac{d}{dt} (\mathbf{I}) = \mathbf{0}$ , it follows that  $\boldsymbol{\sigma}^* = \mathbf{Q} \boldsymbol{\sigma} \mathbf{Q}^T$ . Since (8.52) is satisfied, it follows that the constitutive law is consistent with the Principle of Frame-Indifference.  $\blacksquare$

*Example 4* Is it possible to have  $\boldsymbol{\sigma} = \alpha \mathbf{I}$ , where  $\alpha$  is a scalar that depends on  $\nabla \mathbf{u}$ ? First note that, as required,  $\boldsymbol{\sigma}$  is symmetric. To be consistent with the Principle of Frame-Indifference, it is required that  $\alpha((\nabla \mathbf{u})^*) = \alpha(\nabla \mathbf{u})$ . From Exercise 8.15(d),  $(\nabla \mathbf{u})^* = \mathbf{I} - \mathbf{Q}_0 (\mathbf{I} - \nabla \mathbf{u}) \mathbf{Q}^T$ , or equivalently,

$$\mathbf{I} - (\nabla \mathbf{u})^* = \mathbf{Q}_0 (\mathbf{I} - \nabla \mathbf{u}) \mathbf{Q}^T.$$

One possibility is to use the determinant, since

$$\begin{aligned} \det[\mathbf{I} - (\nabla \mathbf{u})^*] &= \det(\mathbf{Q}_0) \det(\mathbf{I} - \nabla \mathbf{u}) \det(\mathbf{Q}^T) \\ &= \det(\mathbf{I} - \nabla \mathbf{u}). \end{aligned}$$

Consequently, the assumption that  $\boldsymbol{\sigma} = \alpha \mathbf{I}$ , where  $\alpha$  is a scalar function of  $\gamma = \det(\mathbf{I} - \nabla \mathbf{u})$  is consistent with the conditions we have imposed on our constitutive law for the stress.  $\blacksquare$

A consequence of material frame-indifference is that the material must be isotropic, which means that the material properties are the same in all directions. It is possible to generalize the formulation and include non-isotropic materials, and an introduction to this can be found in Batra (2002) and Xiao et al. (2006).

It needs to be pointed out that Frame-Indifference is only imposed on the constitutive law for the stress, and not imposed on the equations of motion. As shown in Exercise 8.15, the momentum equation is not form invariant. More precisely, it is not invariant for Euclidean transformations, but it is for Galilean transformations. This fact is one of the reasons for the rather pointed controversies that surround the subject, and this will be discussed in more detail later.

### 8.9.1 Representation Theorem and Invariants

The assumptions that the stress is objective, and the constitutive law is form invariant, have some interesting consequences. One, which will play an important role in the constitutive modeling, is the next result, known as the Rivlin-Ericksen representation theorem.

**Rivlin-Ericksen Representation Theorem.** *Assume  $\boldsymbol{\sigma}$  and  $\mathbf{R}$  are both symmetric and objective tensors. If  $\boldsymbol{\sigma} = \mathbf{G}(\mathbf{R})$  is form invariant, then the constitutive law can be rewritten as*

$$\boldsymbol{\sigma} = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{R} + \alpha_2 \mathbf{R}^2, \quad (8.58)$$

where the coefficients  $\alpha_0$ ,  $\alpha_1$ , and  $\alpha_2$  are functions of

$$I_R = \text{tr}(\mathbf{R}), \quad (8.59)$$

$$II_R = \frac{1}{2} \left( \text{tr}(\mathbf{R})^2 - \text{tr}(\mathbf{R}^2) \right), \quad (8.60)$$

$$III_R = \det(\mathbf{R}). \quad (8.61)$$

In the above expressions,  $\text{tr}()$  is the trace, and  $\det()$  is the determinant. A proof, with an additional assumption on the constitutive law, is given in Exercise 8.32. The proof for the general case can be found in Rivlin and Ericksen (1955) or Wang (1970).

The three quantities  $I_R$ ,  $II_R$ , and  $III_R$  are the principal invariants of  $\mathbf{R}$ . They are called invariants because their values do not change when changing coordinates using rigid body motion. The proof of this statement is based on the identities  $\text{tr}(\mathbf{RS}) = \text{tr}(\mathbf{SR})$  and  $\det(\mathbf{RS}) = \det(\mathbf{SR})$ . For example,  $III_{R^*} = \det(\mathbf{R}^*) = \det(\mathbf{QRQ}^T) = \det(\mathbf{Q}^T \mathbf{QR}) = \det(\mathbf{R}) = III_R$ . To take advantage of this observation, recall that a symmetric matrix can be diagonalized, and the diagonal entries are the eigenvalues  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$ . With this we have that

$$I_R = \lambda_1 + \lambda_2 + \lambda_3, \quad (8.62)$$

$$II_R = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3, \quad (8.63)$$

$$III_R = \lambda_1\lambda_2\lambda_3. \quad (8.64)$$

This shows that in terms of their dependence on the eigenvalues,  $I_R$  is linear,  $II_R$  is quadratic, and  $III_R$  is cubic. It also shows that the three invariants are independent, in the sense that it is not possible to write one of them in terms of the other two. Some of the properties of the principal invariants are developed in Exercise 8.16.

The ideas underlying the Principle of Frame-Indifference are almost universally accepted by those working in continuum mechanics, and the development used here was adapted from the work of Svendsen and Bertram (1999). This does not mean that all issues related to this principle have been worked out. Most of the attention on this topic is not germane to this textbook, but it is worth providing a glimpse into some of the questions that have arisen. Constitutive laws are macroscopic functions representing the accumulated actions taking place on the atomic scale. This viewpoint was used in Chap. 6 to help explain how the atomic structure of a solid can give rise to the assumed linear law of elasticity. The idea is that it should be possible to derive the constitutive law from more fundamental principles, such as that arise in statistical mechanics. This very attractive proposal brings with it a problem, which is that the Newtonian laws of microscopic physics generally do not satisfy the Principle of Frame-Indifference. Given this then how can one expect that the resulting macroscopic constitutive laws obey this principle? The resolution of this problem involves a closer look at the limit taken when moving from the microscopic to macroscopic scale, and this is discussed in the papers by Speziale (1987) and Murdoch (2006). More general reviews on Frame-Indifference can be found in Speziale (1998) and Frewer (2009).

## 8.10 Newtonian Fluid

To apply the above theory to the study of fluid motion, the first question to consider is, what exactly is a fluid? It is certainly easy to list specific examples. This includes liquids, such as water and mercury at room temperature, as well as gases, such as air. Two of the central characteristics of fluids is their ability to flow, and their inability to retain a specific shape of their own. The important question for us is, how do we translate these observations into a mathematical formula for the stress?

### 8.10.1 Pressure

In developing the constitutive law for a solid, one of the first experiments we considered was what happens when the material is compressed. We found that the displacement in such situations was not constant, and this gave rise to the concept of a strain in an elastic solid. The reason a strain is possible within the solid is indicated in Fig. 6.10. The forces holding the atoms in the lattice enable the solid to support a variable displacement. This is not possible in a fluid. The reason is that the fluid atoms are farther apart, and are able to move past each other with little difficulty. Under compression they will get closer together, and assuming the fluid has come to rest, they are all approximately equidistant from each other. This is the situation, for example, that occurs after you have blown up a balloon. The compression does introduce a stress in the fluid, and it is the same in all directions. This is the concept underlying a pressure, and the resulting constitutive law is

$$\boldsymbol{\sigma} = -p\mathbf{I}, \quad (8.65)$$

where  $\mathbf{I}$  is the identity tensor and  $p$  is the pressure.

### 8.10.2 Viscous Stress

What about the stress when the fluid is moving? A hint on how to answer this is obtained from the usual explanation for how to account for air resistance. When modeling the motion of an object in a fluid, such as a ball falling through the air, it is usually assumed that the drag force is proportional to the velocity. The correct way to say this is that it is proportional to the relative velocity between the fluid and object. The idea is that when atoms of the fluid move together, in parallel, there is no relative velocity and therefore no resistance. It is when the atoms move past each other that the resistance force is generated. The resulting constitutive assumption is that the fluid stress depends on the spatial derivative of the fluid velocity. To get an idea of what this entails, all of the various spatial derivatives of  $\mathbf{v}$  are collected together in the velocity gradient tensor  $\nabla\mathbf{v}$ , given in (8.25). The constitutive assumption is, therefore, that each of the six elements of the stress tensor are functions of the nine derivatives in the velocity gradient. This is not a very appealing result. For example, even if we try to make things easy and assume that the dependence is linear we end up with 54 parameters.

To derive a more manageable theory for fluids, we first use  $\nabla \mathbf{v}$  to introduce two associated tensors. One is the *rate of deformation tensor*, defined as

$$\mathbf{D} \equiv \frac{1}{2} (\nabla \mathbf{v} + \nabla \mathbf{v}^T) = \begin{pmatrix} \frac{\partial v_1}{\partial x} & \frac{1}{2} \left( \frac{\partial v_1}{\partial y} + \frac{\partial v_2}{\partial x} \right) & \frac{1}{2} \left( \frac{\partial v_1}{\partial z} + \frac{\partial v_3}{\partial x} \right) \\ \frac{1}{2} \left( \frac{\partial v_1}{\partial y} + \frac{\partial v_2}{\partial x} \right) & \frac{\partial v_2}{\partial y} & \frac{1}{2} \left( \frac{\partial v_2}{\partial z} + \frac{\partial v_3}{\partial y} \right) \\ \frac{1}{2} \left( \frac{\partial v_1}{\partial z} + \frac{\partial v_3}{\partial x} \right) & \frac{1}{2} \left( \frac{\partial v_2}{\partial z} + \frac{\partial v_3}{\partial y} \right) & \frac{\partial v_3}{\partial z} \end{pmatrix}, \quad (8.66)$$

and the other is the vorticity, or spin, tensor

$$\mathbf{W} \equiv \frac{1}{2} (\nabla \mathbf{v} - \nabla \mathbf{v}^T) = \begin{pmatrix} 0 & \frac{1}{2} \left( \frac{\partial v_1}{\partial y} - \frac{\partial v_2}{\partial x} \right) & \frac{1}{2} \left( \frac{\partial v_1}{\partial z} - \frac{\partial v_3}{\partial x} \right) \\ \frac{1}{2} \left( \frac{\partial v_2}{\partial x} - \frac{\partial v_1}{\partial y} \right) & 0 & \frac{1}{2} \left( \frac{\partial v_2}{\partial z} - \frac{\partial v_3}{\partial y} \right) \\ \frac{1}{2} \left( \frac{\partial v_3}{\partial x} - \frac{\partial v_1}{\partial z} \right) & \frac{1}{2} \left( \frac{\partial v_3}{\partial y} - \frac{\partial v_2}{\partial z} \right) & 0 \end{pmatrix}. \quad (8.67)$$

The most obvious properties of these two tensors are that  $\mathbf{D}$  is symmetric,  $\mathbf{W}$  is skew-symmetric or antisymmetric, and  $\nabla \mathbf{v} = \mathbf{D} + \mathbf{W}$ .

One approach for formulating a constitutive law for the viscous stress is to assume that the dependence on  $\nabla \mathbf{v}$  is linear. Given the requirement that the stress is symmetric, a possible assumption is that

$$\boldsymbol{\sigma} = -p\mathbf{I} + \lambda(\nabla \cdot \mathbf{v})\mathbf{I} + 2\mu\mathbf{D}, \quad (8.68)$$

where  $\lambda$  and  $\mu$  are constants. As it turns out, this is the law used for what is known as a compressible Newtonian fluid. However, although the approach used to obtain this result is almost effortless, it is likely that you would miss the  $\lambda(\nabla \cdot \mathbf{v})\mathbf{I}$  term without some thought as to what “linear and symmetric” means.

In what follows, a systematic reduction is carried out to derive (8.68). The steps involved are made to better understand the physical assumptions underlying the formulation of the constitutive law for the stress. They also provide a segue into how to formulate a constitutive law for other types of fluids, such as the nonlinear theory for what are called Reiner-Rivlin fluids and also power-law fluids. The latter will be considered, briefly, in the next chapter.

### 8.10.2.1 Reduction of the Viscous Stress Function

In its general form, the fluid stress is assumed to not depend on  $\nabla \mathbf{v}$  but, rather, on  $\mathbf{D}$  and  $\mathbf{W}$ . Specifically, the assumption is that

$$\boldsymbol{\sigma} = -p\mathbf{I} + \mathbf{G}(\mathbf{D}, \mathbf{W}), \quad (8.69)$$

where  $\mathbf{G}(\mathbf{0}, \mathbf{0}) = \mathbf{0}$ . In what follows, the specific form of the function  $\mathbf{G}$  is reduced, using the properties of  $\boldsymbol{\sigma}$  and additional simplifying assumptions. Before doing this, note that we have assumed that  $\mathbf{G}$  does not depend explicitly on  $\mathbf{x}$ . This means we are assuming that the fluid is homogeneous, so the constitutive law for the stress does not depend explicitly on position.

*Simplification 1:*  $\boldsymbol{\sigma} = -p\mathbf{I} + \mathbf{G}(\mathbf{D})$

The conclusion that the stress does not depend on  $\mathbf{W}$  is not too surprising because  $\boldsymbol{\sigma}$  is symmetric while  $\mathbf{W}$  is skew-symmetric. The proof, however, comes from the Principle of Material Frame-Indifference. For the rigid body motion given in (8.50), it is shown in Exercise 8.15 that  $\mathbf{D}^* = \mathbf{QDQ}^T$  and  $\mathbf{W}^* = \mathbf{QWQ}^T + \boldsymbol{\Omega}$ , where  $\boldsymbol{\Omega} = \mathbf{Q}'\mathbf{Q}^T$  is skew-symmetric. Therefore, from (8.52) and (8.53), it is required that

$$\mathbf{QG}(\mathbf{D}, \mathbf{W})\mathbf{Q}^T = \mathbf{G}(\mathbf{QDQ}^T, \mathbf{QWQ}^T + \boldsymbol{\Omega}). \quad (8.70)$$

To make our point we do not need to consider every rotation, and it is enough to consider those where  $\mathbf{Q}(0) = \mathbf{I}$  and  $\mathbf{Q}'(0) = \mathbf{M}$  is an arbitrary skew-symmetric matrix. With this, and setting  $t = 0$  in (8.70), it follows that

$$\mathbf{G}(\mathbf{D}, \mathbf{W}) = \mathbf{G}(\mathbf{D}, \mathbf{W} + \mathbf{M}),$$

for every skew-symmetric matrix  $\mathbf{M}$ . The only function capable of this is one that does not depend on  $\mathbf{W}$ . The dependence of  $\mathbf{G}$  on  $\mathbf{D}$  is left open, other than it must satisfy  $\mathbf{QG}(\mathbf{D})\mathbf{Q}^T = \mathbf{G}(\mathbf{QDQ}^T)$ , for all rotations  $\mathbf{Q}$ .

*Simplification 2:*  $\mathbf{G}(\mathbf{D}) = \alpha_0\mathbf{I} + \alpha_1\mathbf{D} + \alpha_2\mathbf{D}^2$

This result is an immediate consequence of the Rivlin-Ericksen theorem (8.58), because both  $\boldsymbol{\sigma}$  and  $\mathbf{D}$  are symmetric and objective (see Exercise 8.15). In this case, the coefficients  $\alpha_0$ ,  $\alpha_1$ , and  $\alpha_2$  have the dimensions of stress and, with one exception, are arbitrary functions of the three principal invariants

$$\text{I}_D = \text{tr}(\mathbf{D}),$$

$$\text{II}_D = \frac{1}{2} \left( \text{tr}(\mathbf{D})^2 - \text{tr}(\mathbf{D}^2) \right),$$

$$\text{III}_D = \det(\mathbf{D}).$$

The exception comes from the requirement that  $\mathbf{G} = \mathbf{0}$  if  $\mathbf{D} = \mathbf{0}$ , which means that  $\alpha_0 = 0$  if  $\mathbf{D} = \mathbf{0}$ .



*Simplification 3:*  $\mathbf{G}(\mathbf{D}) = \lambda I_D \mathbf{I} + 2\mu \mathbf{D}$

This simplification is an assumption, and specifically it is assumed that the stress is a linear function of  $\mathbf{D}$ , with  $\mathbf{G}(\mathbf{0}) = \mathbf{0}$ . This means, using the result of Simplification 2, that  $\alpha_2 = 0$  and  $\alpha_1 = 2\mu$  is a constant. The coefficient  $\alpha_0$  can be a linear function of the elements of  $\mathbf{D}$ . As is evident in how the invariants depend on the eigenvalues, as given in (8.62)–(8.64), only  $I_D$  is linear in  $\mathbf{D}$ . Therefore, it follows that  $\alpha_0 = \lambda I_D$ , where  $\lambda$  is a constant.

## 8.11 Equations of Motion for a Viscous Fluid

The conclusion from the previous section is that the constitutive law for a viscous fluid is

$$\boldsymbol{\sigma} = -p\mathbf{I} + \lambda(\nabla \cdot \mathbf{v})\mathbf{I} + 2\mu\mathbf{D}, \quad (8.71)$$

where  $p$  is the pressure and  $\mathbf{D}$  is the rate of deformation tensor given in (8.66). This is the constitutive law for what is called a *Newtonian fluid*, which means that the stress is a linear function of the rate of deformation tensor. The coefficients  $\lambda$  and  $\mu$  are viscosity parameters. In the engineering literature  $\mu$  is called the dynamic viscosity, or sometimes the shear viscosity. The coefficient  $\lambda$  is known as the second coefficient of viscosity. It can be shown that for the second law of thermodynamics to be satisfied it is required that  $\mu \geq 0$  and  $3\lambda + 2\mu \geq 0$ .

With this, one finds that

$$\nabla \cdot \boldsymbol{\sigma} = -\nabla p + \lambda \nabla(\nabla \cdot \mathbf{v}) + \mu \left( \nabla(\nabla \cdot \mathbf{v}) + \nabla^2 \mathbf{v} \right), \quad (8.72)$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad (8.73)$$

is the Laplacian. The resulting equations of motion are

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0, \quad (8.74)$$

$$\rho \frac{D\mathbf{v}}{Dt} = -\nabla p + (\lambda + \mu) \nabla(\nabla \cdot \mathbf{v}) + \mu \nabla^2 \mathbf{v} + \rho \mathbf{f}. \quad (8.75)$$

The above momentum equation is known as the *Navier-Stokes equation*.

Looking at the equations in (8.74) and (8.75) you soon realize that something is missing. Namely, there are four equations, but five unknowns ( $\rho$ ,  $p$ , and  $\mathbf{v}$ ). What is needed is a second constitutive law, an equation of state, that relates the pressure

and density. Commonly used examples are the ideal gas law  $p = \rho RT$ , and the van der Waals equation

$$p = \rho RT \left( \frac{1}{1 - \beta \rho} - \frac{\alpha \rho}{RT} \right).$$

Both of these expressions contain the temperature  $T$ . For isothermal flows this is assumed constant, but if this is not the case, then it is necessary to derive a balance law for the energy. This was considered earlier, in Sect. 6.10.2, for one-dimensional motion, but will not be considered here.

A few comments are in order related to the constitutive law in (8.71). First, it is important to note that a Newtonian fluid is an assumption of material linearity, and not an assumption of geometric linearity. The domains over which the Newtonian fluid equations are applicable can be, and are routinely, highly variable. A second comment is that there are different ways to derive (8.71). As an example of a different approach, it is possible to obtain (8.71) using the Principle of Dissipation. This was the method used in Sect. 6.10.2, to derive the one-dimensional version of (8.71). The last comment concerns the viscosity coefficients. The dynamic viscosity  $\mu$  is easily measured, and it is not hard to find extensive tables listing its values for a wide range of fluids. This is not the case for the second coefficient of viscosity  $\lambda$ . Even though Newtonian fluids have been studied for almost two centuries, there have been few accurate measurements of  $\lambda$ . There have been analytical derivations of its value, and the most prominent of these is the result that for a dilute monatomic gas  $\lambda = -\frac{2}{3}\mu$ . Using this particular value is known in fluid dynamics as the Stokes hypothesis. In contrast, for a dilute polyatomic gas, it has been reported that the value of  $\lambda$  can be  $100\mu$  or larger (Cramer, 2012), while using what is known as a smoothed-particle hydrodynamics model it has been concluded that  $\lambda = \mu$  (Colagrossi et al., 2017). To help add to the confusion, recent experimental studies suggest that there is no meaningful correlation between  $\lambda$  and  $\mu$  (Dukhin and Goetz, 2009). Fortunately, for a fluid that is incompressible, the value of  $\lambda$  is not needed, and this brings us to the next topic.

### 8.11.1 Incompressibility

If the fluid is assumed to be incompressible, then the constitutive law for the stress is

$$\boldsymbol{\sigma} = -p\mathbf{I} + 2\mu\mathbf{D}. \quad (8.76)$$

The equations of motion in this case reduce to the following:

$$\rho \frac{D\mathbf{v}}{Dt} = -\nabla p + \mu \nabla^2 \mathbf{v} + \rho \mathbf{f}, \quad (8.77)$$

$$\nabla \cdot \mathbf{v} = 0. \quad (8.78)$$

Assuming that the initial density is constant, then  $\rho$  is known, and it is constant. In this case the number of equations matches the number of unknowns ( $p$  and  $\mathbf{v}$ ), and so an equation of state is not needed. Also, these equations appear to be somewhat simpler than the compressible versions in (8.74) and (8.75). Although this may be true, both versions are formidable mathematical problems.

Given our interest in solving (8.77) and (8.78), it is worth spending a moment considering what sort of mathematical problem we are facing. In terms of the velocity, (8.77) is a first-order equation in time and a second-order equation in space. In this sense it is the same as the diffusion equations studied in Chap. 4, and it should not be surprising to find that the kinematic viscosity  $\nu = \mu/\rho$  has the same dimensions as a diffusion coefficient. One of the distinctive differences from a diffusion equation is the nonlinear term  $(\mathbf{v} \cdot \nabla)\mathbf{v}$  hiding in the material derivative. This term is the type of nonlinearity we studied in traffic flow. The nonlinearity, however, means that transform methods, both Fourier and Laplace, will not work. Also, the presence of the viscosity term means that the method of characteristics will not work. Aside from a numerical solution, this leaves similarity methods, perturbation methods, and guessing. We will make heavy use of guessing, and this requires a well-formulated mathematical problem, which means that we need boundary conditions.

Before moving on, it is worth commenting on the earlier statement about the formidability of solving the Navier-Stokes equation. Finding the solution is considered to be one of the greatest unsolved mathematical problems of our time, and this is the reason that it was included as one of the Millennium Prize Problems (Carlson et al. 2006). The person, or team, that first solves this problem will be awarded \$1,000,000 (US).

### 8.11.2 *Boundary and Initial Conditions*

To be able to find the solution of the fluid equations it is necessary to know the boundary conditions. Of interest here are the conditions at a solid boundary, which for us will usually be the container holding the fluid. Although it is often the case that such boundaries are stationary, we will include the possibility that it moves. An example of such a situation arises with a water balloon, where the boundary, the balloon, does not just move, it also deforms.

In the following, let  $S$  be the boundary surface, and assume that the points on the boundary have a known velocity  $\mathbf{v}_S(\mathbf{x}, t)$ .

#### *Impermeability Condition*

The boundary is solid, and this means that the fluid cannot flow through it. To translate this into a boundary condition, let  $\mathbf{n}$  be the unit outward normal to the surface. With this,  $\mathbf{v} \cdot \mathbf{n}$  is the velocity of the fluid in the normal direction, and  $\mathbf{v}_S \cdot \mathbf{n}$  is the velocity of the surface in the normal direction. The boundary condition is one

of continuity, namely that on  $S$  the normal velocity of the fluid is equal to the normal velocity of the surface. Therefore, the mathematical consequence of impermeability is that

$$\mathbf{v} \cdot \mathbf{n} = \mathbf{v}_s \cdot \mathbf{n} \quad \text{on } S. \quad (8.79)$$

If the fluid is incompressible, and  $S$  is the boundary of a bounded domain, then  $\mathbf{v}_s$  must be consistent with the incompressibility assumption (see Exercise 8.20). Also, it should be pointed out that it is implicitly assumed here that the fluid does not separate from the boundary. There are situations where separation occurs, such as in cavitation, and this often results in a very challenging mathematical problem. For obvious reasons, we will avoid such situations in this introductory presentation.

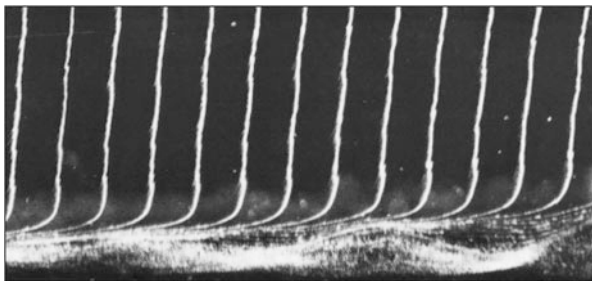
### *No-Slip Condition*

Because of the viscosity, it is assumed that the fluid sticks to the boundary. This means that the fluid velocity, on the boundary, equals the velocity of the boundary. The corresponding boundary condition is

$$\mathbf{v} = \mathbf{v}_s \quad \text{on } S. \quad (8.80)$$

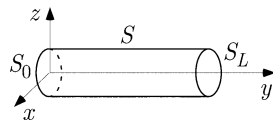
This is known as the *no-slip condition*. It means that not only the normal velocities are equal, as required by the impermeability condition (8.79), it also means that the tangential velocities are equal. In many fluid problems the boundary does not move, and in those cases the no-slip condition is simply  $\mathbf{v} = \mathbf{0}$  on  $S$ . An illustration of this situation is shown in Fig. 8.6.

It is common to have situations where the pressure is prescribed on the boundary. Rather than attempt to write down a general formulation of such situations, it is informative to consider an example.



**Fig. 8.6** Fluid flow over a flat plate illustrating the no-slip boundary condition. The fluid is moving from left to right, and the plate is at the bottom and is not moving. The white curves are indicators of the fluid particles moving with the flow, which show a rapid transition from zero velocity on the plate, to the constant velocity in the upper region

**Fig. 8.7** Geometry of pipe used in the boundary condition example



*Example (Flow in a Pipe)* Most people, when attempting to drink using a straw, create a pressure difference between the two ends of the straw. This is the same idea used to have water flow through a hose or pipe. To formulate the boundary conditions for such a situation, consider the straight pipe shown in Fig. 8.7. The pipe is fixed, and water is flowing through the pipe due to a pressure difference between the two ends. From the no-slip condition we have that  $\mathbf{v} = \mathbf{0}$  on  $S$ , which is the wall of the pipe. At the two ends,  $S_0$  and  $S_L$ , inflow/outflow conditions are used. This means that instead of prescribing the three components of the velocity, we will prescribe its two tangential components and the pressure. Letting  $\mathbf{v} = (u, v, w)$ , then on  $S_0$  we take  $u = w = 0$  and  $p = p_0$ , while at  $S_L$  we take  $u = w = 0$  and  $p = p_L$ . These boundary conditions, along with the equations of motion, form what is known as the Poiseuille flow problem, and the solution in the case of steady flow is derived in Sect. 9.1.2. ■

Another boundary condition that often arises involves the surface between two fluids, such as the interface between water and air. An example of this type of condition is given in the next chapter when studying water waves.

#### Initial Condition

The usual initial condition used for incompressible fluid problems is simply

$$\mathbf{v}(\mathbf{x}, 0) = \mathbf{v}_0(\mathbf{x}), \quad (8.81)$$

where  $\mathbf{v}_0(\mathbf{x})$  is given. Not just any function can be used here. In particular, it must be consistent with incompressibility, and therefore it is required that  $\mathbf{v}_0(\mathbf{x})$  satisfies (8.78). A consistency requirement can also come from the boundary conditions. For example, at a solid surface the impermeability condition (8.79) must be satisfied. This means that  $\mathbf{v}_0 \cdot \mathbf{n} = \mathbf{v}_s(\mathbf{x}, 0) \cdot \mathbf{n}$  on  $S$ . To obtain a well-posed mathematical problem it is not necessary that the tangential components of  $\mathbf{v}_0(\mathbf{x})$  satisfy the no-slip condition. A more complete discussion of the various boundary and initial conditions that can be used to obtain a well-posed mathematical problem involving the Navier-Stokes equation can be found in Temam (2001), and from a more computational viewpoint in Nordström (2017).

Boundary conditions are of supreme importance in the formulation of any physical problem. This is brought up because the equations of motion have been derived from fundamental physical principles. The boundary conditions, in contrast, give the appearance of being tacked on using plausibility arguments why they should be used. The no-slip condition is an example of this. Make no mistake, it is almost universally used for viscous fluid problems. However, as a budding applied mathematician, you should be skeptical of this situation. One question

you might ask is, can it be derived from more fundamental physical principles? Even more important, are there situations where it should not be used? These are difficult questions, and in the early development of fluid dynamics they were controversial topics. Eventually, based on the available experimental evidence, the no-slip condition became the accepted requirement. These questions, however, have started to be asked again. This is due to better experimental methods, and the application of the Navier-Stokes equations to small-scale systems where the no-slip condition is questionable. It should be pointed out that the questions apply to the tangential component of the no-slip condition. The normal component, which is the impermeability condition (8.80), can be derived from the continuity equation (Hutter and Johnk 2004), and is not in question. Those interested in the no-slip condition, and its limitations, should consult the review by Lauga et al. (2007).

## 8.12 Material Equations of Motion

Because elastic solids have the ability to hold their shape, they have a natural reference configuration. For this reason, the material coordinate system is more often used in elasticity. There are a couple of options here for how to determine the material version of the equations of motion. One is to use the chain rule, and convert the spatial derivatives in (8.47) and (8.48) into material derivatives. This is rather tedious, and not very enlightening. Another approach is to derive the equations from the material form of the balance laws, and this is the one used here.

The more interesting equation to derive is the one for linear momentum. To state this result, let  $\bar{B}$  be a volume of material points with boundary surface  $\partial\bar{B}$ . When using spatial coordinates this volume was designated as  $R(0)$ . Using the balance of linear momentum,

$$\frac{d}{dt} \iiint_{\bar{B}} R \mathbf{V} dV_X = \iint_{\partial\bar{B}} \bar{\mathbf{t}} dS_X + \iiint_{\bar{B}} R \bar{\mathbf{F}} dV_X. \quad (8.82)$$

This equation is simply the material version of (8.37). In this equation  $R(\mathbf{X}, t)$  is the density in material coordinates,  $\mathbf{V}(\mathbf{X}, t)$  is the velocity,  $\bar{\mathbf{t}}(\mathbf{X}, t)$  is the force, per unit area, on  $\bar{B}$  due the material exterior to  $\bar{B}$ , and  $\bar{\mathbf{F}}(\mathbf{X}, t)$  is the external body force, per unit mass, in material coordinates. The subscript  $X$  on the volume and surface elements is to indicate that the integration is with respect to material coordinates. For example, if  $dV = dx dy dz$ , then  $dV_X = dX dY dZ$ .

The Cauchy Stress Theorem, given in Sect. 8.6.1, is applicable here with the appropriate changes for the material coordinate system being used. The conclusion is that

$$\bar{\mathbf{t}} = \mathbf{P} \bar{\mathbf{n}}, \quad (8.83)$$

where

$$\mathbf{P} = \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix} \quad (8.84)$$

is known as the *first Piola-Kirchhoff stress tensor* and  $\bar{\mathbf{n}}$  is the unit outward normal to  $\partial B$ . Substituting this into (8.82), and using the fact that this holds for all material volumes  $B$  we conclude that

$$R_0 \frac{\partial^2 \mathbf{U}}{\partial t^2} = \nabla_X \cdot \mathbf{P} + R_0 \bar{\mathbf{F}}. \quad (8.85)$$

This is the momentum equation in material coordinates. The  $\nabla_X \cdot \mathbf{P}$  term is similar to (8.44), except the derivatives are with respect to the elements of  $\mathbf{X}$  instead of  $\mathbf{x}$ . In simplifying the above result the material version of the continuity equation has been used, which is

$$R(\mathbf{X}, t) = \frac{R_0}{\det(\mathbf{F})}, \quad (8.86)$$

where  $\mathbf{F}$  is given in (8.10) and  $R_0 = R(\mathbf{X}, 0)$  is the initial value for the density. Finally, the angular momentum equation in material coordinates is

$$\mathbf{P}\mathbf{F}^T = \mathbf{F}\mathbf{P}^T. \quad (8.87)$$

The derivation of this result, and the continuity equation, is left as an exercise.

The derivation of the linear momentum equation looks to be a replay of the spatial coordinate analysis, and this is correct. What is left is the more interesting step, and that is to relate the stress tensors  $\boldsymbol{\sigma}$  and  $\mathbf{P}$ . Because they consist of the stresses on three orthogonal coordinate surfaces, and the material is undergoing deformation,  $\boldsymbol{\sigma}$  and  $\mathbf{P}$  are not necessarily equal. What we will find is that

$$\mathbf{P} = \det(\mathbf{F})\boldsymbol{\sigma}\mathbf{F}^{-T}. \quad (8.88)$$

To explain where this formula comes from, recall that the stress is force per area. The value of the force in the material and spatial systems is the same, but the area is not. So, if  $\mathbf{F}_0$  is the force, and the areas used in the material and spatial systems are  $\Delta A$  and  $\Delta S$ , respectively, then  $\mathbf{P}$  would be determined using the ratio  $\mathbf{F}_0/\Delta A$  while  $\boldsymbol{\sigma}$  would use  $\mathbf{F}_0/\Delta S$ . What is shown below is that  $\Delta A/\Delta S$ , when you account for the three spatial dimensions, is  $\det(\mathbf{F})\mathbf{F}^{-T}$ .

The rest of this section concerns how to derive (8.88), and the key for doing this is the next result.

**Nanson's formula.** *Given a differential spatial surface area  $dS$  with normal  $\mathbf{n}$ , and the corresponding differential material surface area  $dS_X$  with normal  $\bar{\mathbf{n}}$ , then*

$$\mathbf{n}dS = \det(\mathbf{F})\mathbf{F}^{-T}\bar{\mathbf{n}}dS_X, \quad (8.89)$$

To derive (8.89), suppose that given a material point  $\mathbf{X}_0$ , its spatial counterpart is  $\mathbf{x}_0 = \mathcal{X}(\mathbf{X}_0, t)$ . Assuming  $\mathbf{X}_0$  is on the surface  $\partial\bar{B}$ , then  $\mathbf{x}_0$  is on the surface  $\partial R(t)$ . The definition of the stress vector  $\mathbf{t}$  uses the force on a small piece of the tangent plane at  $\mathbf{X}_0$ , and then lets this region shrink to zero. This is the same idea employed earlier when using the tetrahedron, shown in Fig. 8.5, to define the stress vector in spatial coordinates. The difference here is that it is easier to use pieces of the tangent plane shaped as parallelograms instead of triangles. To construct the parallelogram let  $\mathbf{X}_1$  and  $\mathbf{X}_2$  be two points close to  $\mathbf{X}_0$ , and in the tangent plane. This means that  $\mathbf{X}_i = \mathbf{X}_0 + \Delta\mathbf{X}_i$ . The corresponding points in the spatial system are determined using (8.9), and the result is  $\mathbf{x}_i \approx \mathbf{x}_0 + \mathbf{F}\Delta\mathbf{X}_i$ , where  $\mathbf{F}$  is the Jacobian matrix for  $\mathcal{X}$ , evaluated at  $\mathbf{X}_0$ . Now, the cross-product  $(\mathbf{X}_2 - \mathbf{X}_0) \times (\mathbf{X}_1 - \mathbf{X}_0) = \Delta\mathbf{X}_2 \times \Delta\mathbf{X}_1$  determines the normal direction to the surface, and its length gives the area of the parallelogram. To determine the corresponding information for the spatial system, we use the vector  $(\mathbf{x}_2 - \mathbf{x}_0) \times (\mathbf{x}_1 - \mathbf{x}_0) \approx (\mathbf{F}\Delta\mathbf{X}_2) \times (\mathbf{F}\Delta\mathbf{X}_1)$ . We are going to compare the areas in these two coordinate systems, and for this we need the identity

$$\mathbf{G}^T(\mathbf{G}\mathbf{y} \times \mathbf{G}\mathbf{z}) = \det(\mathbf{G})(\mathbf{y} \times \mathbf{z}).$$

Setting  $\Delta\mathbf{x}_i = \mathbf{x}_i - \mathbf{x}_0$ , then we have shown that

$$\begin{aligned} \Delta\mathbf{x}_2 \times \Delta\mathbf{x}_1 &\approx (\mathbf{F}\Delta\mathbf{X}_2) \times (\mathbf{F}\Delta\mathbf{X}_1) \\ &= \det(\mathbf{F})\mathbf{F}^{-T}(\Delta\mathbf{X}_2 \times \Delta\mathbf{X}_1). \end{aligned}$$

Letting the area of the parallelogram in the spatial system be  $\Delta S$ , and letting  $\Delta S_X$  be the area in the material system, then the above equation can be written as  $\Delta S \mathbf{n} \approx (\Delta S_X) \det(\mathbf{F})\mathbf{F}^{-T}\bar{\mathbf{n}}$ . Taking the limit of  $\Delta\mathbf{X}_i \rightarrow \mathbf{0}$  we obtain the formula in (8.89).

What this means is that, when changing from spatial to material coordinates in a surface integral, the following holds:

$$\iint_{\partial R(t)} \boldsymbol{\sigma} \mathbf{n} dS = \iint_{\partial \bar{B}} \det(\mathbf{F}) \boldsymbol{\sigma} \mathbf{F}^{-T} \bar{\mathbf{n}} dS_X.$$

The consequence of this is that expressing  $\boldsymbol{\sigma} \mathbf{n}$  in terms of material coordinates we obtain  $\det(\mathbf{F}) \boldsymbol{\sigma} \mathbf{F}^{-T} \bar{\mathbf{n}}$ . Given that this equals  $\mathbf{P} \bar{\mathbf{n}}$ , for all material regions  $\bar{B}$ , it follows that  $\mathbf{P} = \det(\mathbf{F}) \boldsymbol{\sigma} \mathbf{F}^{-T}$ .



The equations of motion in material coordinates are given in (8.85)–(8.87). There is nothing particularly unusual about this system of equations. For example, as in the one-dimensional version given in Chap. 6, it is not necessary to solve a differential equation to find the density when using material coordinates. This is because of (8.86). The one new twist that arises in the material version is that the stress tensor  $\mathbf{P}$  is not necessarily symmetric, but satisfies (8.87) instead. As will be seen shortly, this nonsymmetry does complicate the formulation of the constitutive law for the stress.

### 8.12.1 Frame-Indifference

Before formulating constitutive laws in material coordinates, it is first necessary to understand how the Principle of Frame-Indifference applies. For one-dimensional motion, studied in Sect. 6.10, a constitutive law in material coordinates is frame-invariant if its spatial counterpart is frame-invariant. This statement also holds for the three-dimensional case we are now studying. However, rather than doing this on a case-by-case basis, it is easier to simply determine the material form of the two tenets that make up the Principle of Frame-Indifference, as given in Sect. 8.9.

The rigid body change of coordinates used for the material version of frame-indifference is given in (8.51). With this, the form invariance assumption, Tenet 2, is unaffected. In particular, if the constitutive law is that  $\mathbf{P} = \mathbf{G}(\mathbf{Z})$ , and if  $\mathbf{P}^*$  and  $\mathbf{Z}^*$  are the  $\mathbf{X}^*$  system versions of  $\mathbf{P}$  and  $\mathbf{Z}$ , then it must be that  $\mathbf{P}^* = \mathbf{G}(\mathbf{Z}^*)$ .

The objectivity condition, Tenet 1, is affected by the change in coordinates. To determine how, recall that  $\boldsymbol{\sigma}^* = \mathbf{Q}\boldsymbol{\sigma}\mathbf{Q}^T$  and  $\mathbf{F}^* = \mathbf{Q}\mathbf{F}\mathbf{Q}_0^T$ . So, from (8.88),

$$\begin{aligned}\mathbf{P}^* &= \det(\mathbf{F}^*)\boldsymbol{\sigma}^*(\mathbf{F}^*)^{-T} \\ &= \det(\mathbf{F})\mathbf{Q}\boldsymbol{\sigma}\mathbf{Q}^T(\mathbf{Q}\mathbf{F}\mathbf{Q}_0^T)^{-T} \\ &= \det(\mathbf{F})\mathbf{Q}\boldsymbol{\sigma}\mathbf{F}^{-T}\mathbf{Q}_0^T \\ &= \mathbf{Q}\mathbf{P}\mathbf{Q}_0^T.\end{aligned}\tag{8.90}$$

Therefore,  $\mathbf{P}$  satisfies the material form of objectivity if  $\mathbf{P}^* = \mathbf{Q}\mathbf{P}\mathbf{Q}_0^T$ , for all rotations  $\mathbf{Q}$ .

### 8.12.2 Elastic Solid

We are now at the point where we need to formulate a constitutive law for the stress. The assumption to be made is that the material is elastic. As you might recall, in Sect. 6.8, this assumption in one dimension means that the stress is assumed to be a function of  $\partial_X U$ . The three-dimensional version of this first derivative is

$$\nabla_X \mathbf{U} = \begin{pmatrix} \frac{\partial U_1}{\partial X} & \frac{\partial U_1}{\partial Y} & \frac{\partial U_1}{\partial Z} \\ \frac{\partial U_2}{\partial X} & \frac{\partial U_2}{\partial Y} & \frac{\partial U_2}{\partial Z} \\ \frac{\partial U_3}{\partial X} & \frac{\partial U_3}{\partial Y} & \frac{\partial U_3}{\partial Z} \end{pmatrix}, \quad (8.91)$$

which is known as the *displacement gradient*. It is related to the deformation gradient, given in (8.10), through the identity  $\nabla_X \mathbf{U} = \mathbf{F} - \mathbf{I}$ . It should be noted that, unlike the case in Sect. 6.8,  $\nabla_X \mathbf{U}$  is not a strain (see Exercise 8.26).

An elastic solid is one for which the stress is a function of  $\nabla_X \mathbf{U}$ . One might guess, generalizing (6.51), that a possible constitutive law is  $\mathbf{P} = \lambda \nabla_X \mathbf{U}$ , where  $\lambda$  is a constant. However, as the next example shows, this does not satisfy the Principle of Frame-Indifference.

*Example* Suppose it is assumed that  $\mathbf{P} = \lambda \nabla_X \mathbf{U}$ . It is not hard to show that  $\mathbf{U}^* = \mathbf{Q}\mathbf{U} + (\mathbf{Q} - \mathbf{Q}_0)\mathbf{X} + \mathbf{b} - \mathbf{b}_0$ , from which it follows that (see Exercise 8.15)

$$\nabla_{X^*} \mathbf{U}^* = \mathbf{Q}(\nabla_X \mathbf{U})\mathbf{Q}_0^T + \mathbf{Q}\mathbf{Q}_0^T - \mathbf{I}. \quad (8.92)$$

Form invariance requires that  $\mathbf{P}^* = \lambda \nabla_{X^*} \mathbf{U}^*$ . Objectivity, as given in (8.90), requires that  $\mathbf{P}^* = \mathbf{Q}\mathbf{P}\mathbf{Q}_0^T$ . Consequently, for the assumed constitutive law, it is required that  $\nabla_{X^*} \mathbf{U}^* = \mathbf{Q}(\nabla_X \mathbf{U})\mathbf{Q}_0^T$ . Given the result in (8.92), this does not hold, and therefore this constitutive law does not satisfy the Principle of Frame-Indifference. ■

To derive a constitutive law for an elastic material it makes things easier if the stress is factored, and written as

$$\mathbf{P} = \mathbf{F}\mathbf{S}, \quad (8.93)$$

where  $\mathbf{S}$  is known as the *second Piola-Kirchhoff stress tensor*. As will be explained below, the requirements on  $\mathbf{S}$  are that

$$\mathbf{S}^T = \mathbf{S}, \quad (8.94)$$

and

$$\mathbf{S}^* = \mathbf{Q}_0 \mathbf{S} \mathbf{Q}_0^T. \quad (8.95)$$

To explain the symmetry requirement (8.94), it follows from (8.88) that  $\mathbf{S} = \det(\mathbf{F})\mathbf{F}^{-1}\boldsymbol{\sigma}\mathbf{F}^{-T}$ . Consequently, since  $\boldsymbol{\sigma}$  is symmetric, it follows that  $\mathbf{S}$  is symmetric. This also means, not unexpectedly, that  $\mathbf{P}$  satisfies the angular momentum condition (8.87). The condition in (8.95) comes directly from (8.90).

It is assumed that  $\mathbf{S}$  depends on  $\mathbf{F}$ , but this dependence is through a quantity  $\mathbf{C}$  that has the same properties as  $\mathbf{S}$ . Specifically,  $\mathbf{C}$  is symmetric and  $\mathbf{C}^* = \mathbf{Q}_0 \mathbf{C} \mathbf{Q}_0^T$ . The one usually used in elasticity is

$$\mathbf{C} = \mathbf{F}^T \mathbf{F}, \quad (8.96)$$

which is known as the *right Cauchy-Green deformation tensor*. The reason for using  $\mathbf{C}$  is based, in part, on the next result.

**Polar Decomposition Theorem.** *Given a matrix  $\mathbf{A}$ , there is a rotation matrix  $\mathbf{Q}$ , along with symmetric matrices  $\mathbf{R}$  and  $\mathbf{L}$ , so that  $\mathbf{A} = \mathbf{Q}\mathbf{R}$  and  $\mathbf{A} = \mathbf{L}\mathbf{Q}$ . If  $\mathbf{A}$  is invertible, then  $\mathbf{R}$  and  $\mathbf{L}$  are positive definite.*

Consequently, we can write  $\mathbf{F} = \mathbf{Q}\mathbf{R}$ . Given that we are considering strain, the rotation  $\mathbf{Q}$  is of no interest. The matrix  $\mathbf{R}$ , on the other hand, is the three-dimensional version of the extension ratio we used in Sect. 6.8 to determine the strain. Determining  $\mathbf{R}$  is straightforward, but it requires some effort. One way to avoid having to do this is to note that  $\mathbf{C} = \mathbf{F}^T \mathbf{F} = \mathbf{R}^2$ .

As in Table 6.4, there are different ways to use the extension ratio to define the strain. The conventional choice is to take

$$\mathbf{E} = \frac{1}{2}(\mathbf{C} - \mathbf{I}), \quad (8.97)$$

which is known as the *Green strain tensor*. So, assuming that  $\mathbf{S}$  depends on  $\mathbf{C}$  is equivalent to assuming it depends on  $\mathbf{E}$ . Why  $\mathbf{E}$  qualifies as a measure of strain is explained below (also, see Exercise 8.26).

The usual assumption is that  $\mathbf{S}$  is a linear function of the strain tensor. Using the Green strain tensor, this means that it is assumed

$$\mathbf{S} = \lambda_s I_E \mathbf{I} + 2\mu_s \mathbf{E}, \quad (8.98)$$

where  $I_E = \text{tr}(\mathbf{E})$ , and  $\lambda_s$  and  $\mu_s$  are constants. The resulting constitutive law for  $\mathbf{P}$  is

$$\mathbf{P} = \lambda_s I_E \mathbf{F} + 2\mu_s \mathbf{F}\mathbf{E}. \quad (8.99)$$

How well this particular constitutive assumption does, in comparison to other possible choices, is considered in Exercise 8.25.

What has been left unexplained is why  $\mathbf{E}$  qualifies as a strain, while  $\mathbf{F}$  and  $\mathbf{C}$  do not. The reason, as outlined in Sect. 6.10.1, is that the stress is determined by the relative motion of the particles in the material. So, for example, with a uniform displacement or for a rigid body rotation, the stress should not change. A strain is a variable that depends on the deformation, but, like the stress, does not change when there is rigid body motion. In particular, to qualify to be a strain measure it is required that the variable be zero for a rigid body motion. To explore this a bit

more, if  $\mathcal{X} = \mathbf{Q}\mathbf{X} + \mathbf{b}$ , then  $\mathbf{F} = \mathbf{Q}$ ,  $\mathbf{C} = \mathbf{I}$ , and  $\mathbf{E} = \mathbf{0}$ . Consequently, the latter is a strain and the other two are not. It should also be mentioned that a strain is implicitly assumed to be monotonic in the sense that the more a material is pulled, the greater the strain (in the direction of the pulling). How to derive mathematical requirements for monotonicity in three-dimensions will not be considered here, but the rigid body motion requirement for a strain is explored in more detail in Exercise 8.26.

### 8.12.3 Linear Elasticity

The constitutive law for the stress  $\mathbf{P}$  in (8.99) is general, in the sense that it only requires the material to be isotropic. We are going to now derive an approximation of it that is based on the assumption of small strains. To begin, since  $\mathbf{F} = \mathbf{I} + \nabla_X \mathbf{U}$ , then

$$\begin{aligned} \mathbf{C} &= (\mathbf{I} + \nabla_X \mathbf{U})^T (\mathbf{I} + \nabla_X \mathbf{U}) \\ &= \mathbf{I} + \nabla_X \mathbf{U} + (\nabla_X \mathbf{U})^T + (\nabla_X \mathbf{U})^T (\nabla_X \mathbf{U}) \\ &\approx \mathbf{I} + \nabla_X \mathbf{U} + (\nabla_X \mathbf{U})^T. \end{aligned} \quad (8.100)$$

With this we have that

$$\begin{aligned} \mathbf{E} &= \frac{1}{2}(\mathbf{C} - \mathbf{I}) \\ &\approx \frac{1}{2}(\nabla_X \mathbf{U} + (\nabla_X \mathbf{U})^T), \end{aligned} \quad (8.101)$$

and so  $\mathbf{I}_E \approx \nabla_X \cdot \mathbf{U}$ . So, from (8.99),

$$\begin{aligned} \mathbf{P} &\approx \lambda_s (\nabla_X \cdot \mathbf{U})(\mathbf{I} + \nabla_X \mathbf{U}) + \mu_s (\mathbf{I} + \nabla_X \mathbf{U})(\nabla_X \mathbf{U} + (\nabla_X \mathbf{U})^T) \\ &\approx \lambda_s (\nabla_X \cdot \mathbf{U})\mathbf{I} + \mu_s (\nabla_X \mathbf{U} + (\nabla_X \mathbf{U})^T). \end{aligned} \quad (8.102)$$

This can be rewritten as

$$\mathbf{P} = \lambda_s (\nabla_X \cdot \mathbf{U})\mathbf{I} + 2\mu_s \mathbf{E}_0, \quad (8.103)$$

where

$$\mathbf{E}_0 = \frac{1}{2}(\nabla_X \mathbf{U} + \nabla_X \mathbf{U}^T). \quad (8.104)$$

This is the constitutive law for linear elasticity, and it is the three-dimensional version of (6.51). The coefficients  $\lambda_s$  and  $\mu_s$  are called the Lamè constants, and

in the engineering literature  $\mu_s$  is referred to as the shear modulus. Also,  $\mathbf{E}_0$  is the linearized Green strain tensor, or what is identified as the Lagrangian strain in Table 6.4.

With (8.103) the equation of motion given in (8.85) reduces to

$$R_0 \frac{\partial^2 \mathbf{U}}{\partial t^2} = (\lambda_s + \mu_s) \nabla_X (\nabla_X \cdot \mathbf{U}) + \mu_s \nabla_X^2 \mathbf{U} + R_0 \bar{\mathbf{F}}. \quad (8.105)$$

This is known as the Navier equation.

There are striking similarities between the constitutive laws, and equations of motion, for fluids and elastic solids. For example, the constitutive law for a linearly elastic solid given in (8.103) is very similar to the fluid law in (8.71), and the Navier equation (8.105) are similar to the Navier-Stokes equation (8.75). There are also differences, and one of the more obvious ones is that elasticity uses the displacement gradient while fluids use the velocity gradient in the formulation of the constitutive law for the stress. However, a perhaps more subtle difference is that the Navier-Stokes equations are obtained using an assumption of material linearity, while the Navier equations require both material and geometric linearity.

## 8.13 Energy Equation

One of the central components in characterizing a mechanical system is the energy. This was discussed in Chap. 6, and in the process a variety of new variables were introduced. A different tack is taken here, and we will derive the energy formulation from what we already know. The key player is the momentum equation (8.48). Taking the dot product with the velocity, and rearranging the terms one obtains the equation

$$\frac{1}{2} \rho \frac{D}{Dt} (\mathbf{v} \cdot \mathbf{v}) = \nabla \cdot (\boldsymbol{\sigma} \mathbf{v}) - \text{tr}(\boldsymbol{\sigma} \mathbf{D}) + \rho \mathbf{v} \cdot \mathbf{f}. \quad (8.106)$$

This is known as the mechanical energy equation. Using the continuity equation (8.47), then (8.106) can be rewritten as

$$\frac{D}{Dt} \left( \frac{1}{2} \rho \mathbf{v} \cdot \mathbf{v} \right) + \left( \frac{1}{2} \rho \mathbf{v} \cdot \mathbf{v} \right) (\nabla \cdot \mathbf{v}) = \nabla \cdot (\boldsymbol{\sigma} \mathbf{v}) - \text{tr}(\boldsymbol{\sigma} \mathbf{D}) + \rho \mathbf{v} \cdot \mathbf{f}. \quad (8.107)$$

This has the form of the general balance law given in (8.31), where  $f = \frac{1}{2} \rho \mathbf{v} \cdot \mathbf{v}$  is the kinetic energy density. To express this in integral form, from (8.30) we have that

$$\frac{d}{dt} \iiint_{R(t)} \frac{1}{2} \rho \mathbf{v} \cdot \mathbf{v} dV = \iint_{\partial R(t)} (\boldsymbol{\sigma} \mathbf{v}) \cdot \mathbf{n} dS + \iiint_{R(t)} \rho \mathbf{v} \cdot \mathbf{f} dV - \iiint_{R(t)} \text{tr}(\boldsymbol{\sigma} \mathbf{D}) dV. \quad (8.108)$$

This shows that the rate of change of the kinetic energy is due to three contributions. The first term on the right, the surface integral, is the rate of work done by surface forces. Given the form of the general balance law in (8.30), this can be interpreted as an energy flux term, with  $\mathbf{J} = -\boldsymbol{\sigma} \mathbf{v}$ . In a similar manner, the integral of  $\rho \mathbf{v} \cdot \mathbf{f}$  is the rate of work done by the body forces. This leaves the integral involving  $\text{tr}(\boldsymbol{\sigma} \mathbf{D})$ . Without the constitutive law for the stress it is not obvious how to interpret this term, and usually in continuum mechanics it is given the rather vague name of the stress power. With this, this term is stated to be rate of work by the stress per unit volume.

What we have in (8.107) and (8.108) are energy balance equations. Usually when considering energy there is a term for the kinetic energy, and another for the potential energy. The kinetic term we have. If there is a contribution of the potential energy it is hidden in either the stress power or the external forcing terms. It is the stress power that is of interest because it is unclear at the moment what this is, and so it is assumed that there are no body forces.

### 8.13.1 Incompressible Viscous Fluid

Assuming that  $\boldsymbol{\sigma} = -p\mathbf{I} + 2\mu\mathbf{D}$ , and  $\nabla \cdot \mathbf{v} = 0$ , then

$$\begin{aligned} \text{tr}(\boldsymbol{\sigma} \mathbf{D}) &= \text{tr}((-p\mathbf{I} + 2\mu\mathbf{D})\mathbf{D}) = -p \text{tr}(\mathbf{D}) + 2\mu \text{tr}(\mathbf{D}\mathbf{D}) \\ &= 2\mu \text{tr}(\mathbf{D}^2). \end{aligned}$$

In fluid mechanics it is conventional to set

$$\Phi = 2\mu \text{tr}(\mathbf{D}^2), \quad (8.109)$$

which is known as the viscous dissipation function. With this, (8.108) becomes

$$\frac{d}{dt} \iiint_{R(t)} \frac{1}{2} \rho \mathbf{v} \cdot \mathbf{v} dV = \iint_{\partial R(t)} (\boldsymbol{\sigma} \mathbf{v}) \cdot \mathbf{n} dS - \iiint_{R(t)} \Phi dV. \quad (8.110)$$

Given that  $\Phi \geq 0$ , then the above expression shows that the kinetic energy decreases due to this term. Physically, this means that the stress power in this particular example accounts for the loss of energy due to viscosity. For this reason, a viscous fluid does not conserve energy. As a final comment, note that for an incompressible viscous fluid, with no body forces, there is no potential energy term in the energy equation.

### 8.13.2 Elasticity

To apply the above arguments to an elastic material we first need to derive the energy equation in material coordinates. Proceeding in a similar manner as before, we take the dot product of the momentum equation (8.85) with the velocity. Remembering our earlier assumption that there are no body forces, the result is

$$\frac{\partial}{\partial t} \left( \frac{1}{2} R_0 \mathbf{V} \cdot \mathbf{V} \right) = \nabla_X \cdot (\mathbf{P}^T \mathbf{V}) - \text{tr}(\mathbf{P}^T \nabla_X \mathbf{V}). \quad (8.111)$$

The above equation states that the rate of change in the kinetic energy density is balanced by the energy flux and the negative of the stress power  $\text{tr}(\mathbf{P}^T \nabla_X \mathbf{V})$ . It is the latter term that needs to be sorted out, and this requires the constitutive law for the stress.

To express the stress power in terms that make its contribution more evident we need to first derive a few identities involving the derivative. So, given a tensor  $\mathbf{R}(t)$ , we have that

$$\begin{aligned} \frac{\partial}{\partial t} (\mathbf{R}^2) &= \frac{\partial}{\partial t} (\mathbf{R}\mathbf{R}) \\ &= \frac{\partial \mathbf{R}}{\partial t} \mathbf{R} + \mathbf{R} \frac{\partial \mathbf{R}}{\partial t}. \end{aligned}$$

This shows the usual power rule does not apply to tensors, but in taking the trace of the above expression we get

$$\frac{\partial}{\partial t} \text{tr}(\mathbf{R}^2) = \text{tr} \left( \mathbf{R} \frac{\partial \mathbf{R}}{\partial t} \right) + \text{tr} \left( \mathbf{R} \frac{\partial \mathbf{R}}{\partial t} \right) = 2 \text{tr} \left( \mathbf{R} \frac{\partial \mathbf{R}}{\partial t} \right). \quad (8.112)$$

In a similar manner one can show that

$$\frac{\partial}{\partial t} \text{tr}(\mathbf{R}^T \mathbf{R}) = \frac{\partial}{\partial t} \text{tr}(\mathbf{R}\mathbf{R}^T) = 2 \text{tr} \left( \mathbf{R}^T \frac{\partial \mathbf{R}}{\partial t} \right).$$

Using the constitutive law in (8.103), and the results from Exercise (8.7), we have that

$$\begin{aligned} \text{tr}(\mathbf{P}^T \nabla_X \mathbf{V}) &= \text{tr} \left( (\lambda_s \mathbf{I}_E \mathbf{F}^T + 2\mu_s \mathbf{E} \mathbf{F}^T) \frac{\partial}{\partial t} \mathbf{F} \right) \\ &= \lambda_s \mathbf{I}_E \text{tr} \left( \mathbf{F}^T \frac{\partial}{\partial t} \mathbf{F} \right) + \mu_s \text{tr} \left( (\mathbf{F}^T \mathbf{F} - \mathbf{I}) \mathbf{F}^T \frac{\partial}{\partial t} \mathbf{F} \right) \\ &= \frac{1}{2} (\lambda_s \mathbf{I}_E - \mu_s) \frac{\partial}{\partial t} \text{tr}(\mathbf{F}^T \mathbf{F}) + \mu_s \text{tr} \left( \mathbf{F}^T \mathbf{F} \mathbf{F}^T \frac{\partial}{\partial t} \mathbf{F} \right). \end{aligned} \quad (8.113)$$

It is convenient at this point to introduce the *left Cauchy-Green deformation tensor*, defined as

$$\mathbf{B} = \mathbf{F}\mathbf{F}^T. \quad (8.114)$$

From (8.112),

$$\begin{aligned} \frac{\partial}{\partial t} \text{tr}(\mathbf{B}^2) &= 2 \text{tr} \left( \mathbf{B} \frac{\partial \mathbf{B}}{\partial t} \right) \\ &= 4 \text{tr} \left( \mathbf{F}^T \mathbf{F} \mathbf{F}^T \frac{\partial}{\partial t} \mathbf{F} \right). \end{aligned}$$

Also, from (8.97),  $I_E = \frac{1}{2} (\text{tr}(\mathbf{F}^T \mathbf{F}) - 3)$ . Substituting these into (8.113) we obtain

$$\begin{aligned} \text{tr}(\mathbf{P} \nabla_X \mathbf{V}) &= \frac{1}{4} \left( \lambda_s \text{tr}(\mathbf{F}^T \mathbf{F}) - 3\lambda_s - 2\mu_s \right) \frac{\partial}{\partial t} \text{tr}(\mathbf{F}^T \mathbf{F}) + \frac{1}{4} \mu_s \frac{\partial}{\partial t} \text{tr}(\mathbf{B}^2) \\ &= \frac{\partial}{\partial t} \left[ \frac{1}{8} \lambda_s M^2 - \frac{1}{4} (3\lambda_s + 2\mu_s) M + \frac{1}{4} \mu_s \text{tr}(\mathbf{B}^2) \right]. \end{aligned} \quad (8.115)$$

where  $M = \text{tr}(\mathbf{F}^T \mathbf{F})$ .

Inserting (8.115) into (8.111), the conclusion is that

$$\frac{\partial}{\partial t} (K + U_p) = \nabla_X \cdot (\mathbf{P}^T \mathbf{V}), \quad (8.116)$$

where

$$K = \frac{1}{2} R_0 \mathbf{V} \cdot \mathbf{V} \quad (8.117)$$

is the kinetic energy density, and

$$U_p = \frac{1}{8} \lambda_s M^2 - \frac{1}{4} (3\lambda_s + 2\mu_s) M + \frac{1}{4} \mu_s \text{tr}(\mathbf{B}^2) + \frac{3}{8} (3\lambda_s + 2\mu_s) \quad (8.118)$$

is the potential energy density relative to the reference state. Using the Green strain tensor (8.97), and the properties of the trace, the above expression can be written as

$$U_p = \frac{1}{2} \lambda_s [\text{tr}(\mathbf{E})]^2 + \mu_s \text{tr}(\mathbf{E}^2). \quad (8.119)$$

The function  $U_p$  is the energy stored in the elastic material due to the deformation. For this reason it is often called the stored energy function, or the strain energy function. One conclusion coming from (8.116) is that the total energy  $K + U_p$  only



changes due to the energy flux  $-\mathbf{P}^T \mathbf{V}$ . It does not change because of a loss due to dissipation, such as happened with a viscous fluid.

## Exercises

### Sections 8.2 and 8.3

**8.1** This exercise is based on the definition of the material and spatial coordinate systems.

- Suppose a particle that started at location  $(2, 0, -1)$  is, at  $t = 4$ , located at  $(1, 0, 1)$ . What is  $\mathbf{X}$  for this particle? For this particle, what is  $\mathbf{U}(\mathbf{X}, 4)$ ? What is the spatial coordinate  $\mathbf{x}$  at  $t = 4$ , for this particle, and what is the corresponding value of  $\mathbf{u}(\mathbf{x}, 4)$ ?
- For another particle,  $\mathbf{u}(\mathbf{x}, 3) = (1, 0, 0)$  for  $\mathbf{x} = (2, 1, 1)$ . What is  $\mathbf{X}$  for this particle?

**8.2** A motion of the form  $x = \alpha(t)X$ ,  $y = Y/\alpha(t)$ ,  $z = Z$ , where  $\alpha(0) = 1$  and  $\alpha > 0$ , is an example of pure shear.

- Give a geometric interpretation of this motion by describing what happens to the unit cube  $0 \leq X \leq 1$ ,  $0 \leq Y \leq 1$ ,  $0 \leq Z \leq 1$ .
- Find  $\mathbf{U}$ ,  $\mathbf{u}$ ,  $\mathbf{V}$ , and  $\mathbf{v}$ .
- Verify that  $\mathbf{v} = \frac{D\mathbf{u}}{Dt}$ .
- Show that  $\nabla \cdot \mathbf{v} = 0$ .

**8.3** Consider the motion  $x = X + \alpha(t)Y$ ,  $y = Y + \alpha(t)Z$ ,  $z = Z$ , where  $\alpha(0) = 0$ .

- Give a geometric interpretation of this motion by describing what happens to the unit cube  $0 \leq X \leq 1$ ,  $0 \leq Y \leq 1$ ,  $0 \leq Z \leq 1$ .
- Find  $\mathbf{U}$ ,  $\mathbf{u}$ ,  $\mathbf{V}$ , and  $\mathbf{v}$ .
- Verify that  $\mathbf{v} = \frac{D\mathbf{u}}{Dt}$ .
- Show that  $\nabla \cdot \mathbf{v} = 0$ .

**8.4** This problem explores some of the consequences of rigid body motion, as defined in (8.13).

- As expressed in (8.13), rigid body motion consists of a rotation followed by a translation. Show that it can also be written as a translation followed by a rotation.
- Find  $\mathbf{U}$ ,  $\mathbf{V}$ , and  $\mathbf{F}$ .
- Show that  $\mathbf{v} = \mathbf{H}\mathbf{x} + \mathbf{h}$ , where  $\mathbf{H}(t) = \mathbf{Q}'\mathbf{Q}^T$  and  $\mathbf{h}(t) = \mathbf{b}' - \mathbf{H}\mathbf{b}$ .

**8.5** Linear flow occurs when  $\mathbf{v} = \mathbf{H}\mathbf{x} + \mathbf{h}$ , where the matrix  $\mathbf{H}$  and vector  $\mathbf{h}$  can depend on  $t$  (but are independent of  $\mathbf{x}$ ).

- Find  $\mathbf{H}$  and  $\mathbf{h}$  for uniform dilatation and for simple shear. The case for rigid body motion is considered in the previous problem.
- Show that linear motion is possible for an incompressible material only if  $\text{tr}(\mathbf{H}) = 0$ .
- Does simple shear satisfy the condition in part (b)?
- Find  $\frac{D}{Dt}\mathbf{v}$  as a function of  $\mathbf{x}$ .

**8.6** For a homogeneous deformation  $\mathcal{X} = \mathbf{G}\mathbf{X} + \mathbf{g}$ , where the matrix  $\mathbf{G}$  and vector  $\mathbf{g}$  can depend on  $t$  (but are independent of  $\mathbf{X}$ ).

- What are  $\mathbf{G}$  and  $\mathbf{g}$  at  $t = 0$ ?
- Find  $\mathbf{G}$  and  $\mathbf{g}$  for uniform dilatation, for simple shear, and for rigid body motion.
- Show that for a homogeneous deformation, all straight lines remain straight lines under the deformation.
- What is the deformation gradient for a homogeneous deformation?
- Show that  $\mathbf{v} = \mathbf{G}'\mathbf{G}^{-1}(\mathbf{x} - \mathbf{g}) + \mathbf{g}'$ .
- Using the result from part (e), show that a homogeneous deformation is a linear flow (see Exercise 8.5).

**8.7** This problem develops some of the formulas relating the displacement and velocity in spatial and material coordinates. The operators  $\nabla_X$  and  $\nabla$  are defined as used in (8.24) and (8.25), respectively.

- Show that  $\mathbf{F} = \mathbf{I} + \nabla_X \mathbf{U}$ .
- Show that  $\partial_t \mathbf{F} = \nabla_X \mathbf{V}$ .
- Let  $\mathbf{G}(\mathbf{X}, t)$  be a vector function in material coordinates and let its spatial version be  $\mathbf{g}(\mathbf{x}, t)$ . This means that  $\mathbf{G}(\mathbf{X}, t) = \mathbf{g}(\mathcal{X}(\mathbf{X}, t), t)$ , and from this equation show that  $\nabla_X \mathbf{G} = (\nabla \mathbf{g})\mathbf{F}$ .
- Show that  $\nabla_X \mathbf{U} = (\mathbf{I} - \nabla \mathbf{u})^{-1}(\nabla \mathbf{u})$  and  $\nabla \mathbf{u} = (\nabla_X \mathbf{U})(\mathbf{I} + \nabla_X \mathbf{U})^{-1}$ .
- Use the result from part (d) to show that in spatial coordinates,  $\mathbf{F} = (\mathbf{I} - \nabla \mathbf{u})^{-1}$ .
- Show that  $\nabla_X \mathbf{V} = (\nabla \mathbf{v})(\mathbf{I} - \nabla \mathbf{u})^{-1}$  and  $\nabla \mathbf{v} = (\nabla_X \mathbf{V})(\mathbf{I} + \nabla_X \mathbf{U})^{-1}$ .

**8.8** This problem develops some of the connections between the displacement and velocity in spatial coordinates.

- Show that to find  $\mathbf{v}$  given  $\mathbf{u}$  one must solve  $(\mathbf{I} - \nabla \mathbf{u})\mathbf{v} = \partial_t \mathbf{u}$ , where  $\mathbf{u} = (u_1, u_2, u_3)$  and

$$\nabla \mathbf{u} = \begin{pmatrix} \frac{\partial u_1}{\partial x} & \frac{\partial u_1}{\partial y} & \frac{\partial u_1}{\partial z} \\ \frac{\partial u_2}{\partial x} & \frac{\partial u_2}{\partial y} & \frac{\partial u_2}{\partial z} \\ \frac{\partial u_3}{\partial x} & \frac{\partial u_3}{\partial y} & \frac{\partial u_3}{\partial z} \end{pmatrix}.$$

- (b) In the case of when the motion is one dimensional, show that the formula in part (a) reduces to (6.14).
- (c) Verify the result from part (a) for uniform dilatation, as given in Sect. 8.2.
- (d) Suppose  $\mathbf{v}$  is known and one wants to determine  $\mathbf{u}$ . Explain why one way this can be done is by solving  $\partial_t \mathbf{u} + (\mathbf{v} \cdot \nabla) \mathbf{u} = \mathbf{v}$ , with  $\mathbf{u}(\mathbf{x}, 0) = \mathbf{0}$ .
- (e) Another method for finding  $\mathbf{u}$  can be derived by reverting to material coordinates. Given a particle that starts out at  $\mathbf{X}$ , explain why the position function  $\bar{\mathbf{X}}(t)$  of that particle satisfies the ordinary differential equation  $\bar{\mathbf{X}}' = \mathbf{v}(\bar{\mathbf{X}}, t)$ , where  $\bar{\mathbf{X}}(0) = \mathbf{X}$ . Given the solution of this problem, assume the equation  $\mathbf{x} = \bar{\mathbf{X}}$  is solved for  $\mathbf{X}$ , to obtain  $\mathbf{X} = \mathbf{x}(\mathbf{x}, t)$ . With this, explain why the displacement is  $\mathbf{u} = \bar{\mathbf{X}} - \mathbf{X}$ , where  $\mathbf{X} = \mathbf{x}(\mathbf{x}, t)$ .
- (f) Suppose that  $\mathbf{v} = (\gamma y, 0, 0)$ , where  $\gamma$  is a constant. Show that  $\mathbf{u} = (\gamma y t, 0, 0)$ . Also, explain why the position of the particle that starts out at  $(x_0, y_0, z_0)$  is  $(x_0 + \gamma y_0 t, y_0, z_0)$ .

## Sections 8.4–8.8

**8.9** If the densities are per unit mass, then the general balance law (8.30) takes the form

$$\frac{d}{dt} \iiint_{R(t)} \rho f(\mathbf{x}, t) dV = - \iint_{\partial R(t)} \mathbf{J} \cdot \mathbf{n} dS + \iiint_{R(t)} \rho Q(\mathbf{x}, t) dV.$$

Show that in this case (8.31) is replaced with

$$\rho \frac{Df}{Dt} = -\nabla \cdot \mathbf{J} + \rho Q.$$

**8.10** Suppose the stress tensor is

$$\boldsymbol{\sigma} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & -1 \\ 3 & -1 & 1 \end{pmatrix}.$$

Consider the unit cube  $0 \leq x \leq 1$ ,  $0 \leq y \leq 1$ , and  $0 \leq z \leq 1$ . Find the stress vector on each face of the cube.

**8.11** Suppose the stress tensor is

$$\boldsymbol{\sigma} = \begin{pmatrix} x & yz & 2 \\ yz & y & x \\ 2 & x & z \end{pmatrix}.$$

- (a) Assuming there are no body forces, explain why it is not possible that the material is at rest.
- (b) What would the body force need to be so the material is at rest?
- (c) Find the stress on the ball  $x^2 + y^2 + z^2 \leq 1$ , at (1)  $(1, 0, 0)$ , (2)  $(0, 1, 0)$ , and (3)  $(0, -1, 0)$ .

**8.12** Suppose it is known that the stress is identically zero, and there are no body forces.

- (a) What is the resulting displacement in material and in spatial coordinates?
- (b) Explain why your conclusion from part (a) holds in the case of when the stress tensor is assumed to be constant.

## Section 8.9

**8.13** Find  $\mathbf{D}$  and then calculate the invariants  $I_D$ ,  $\Pi_D$ ,  $\text{III}_D$ .

- (a) For the motion in Exercise 8.2.
- (b) For the motion in Exercise 8.3.

**8.14** Determine if the following can be used for constitutive laws. Note that the formulas in Exercise 8.15 can be useful here. Also, assume that  $\lambda$  is a positive constant.

- (a)  $\boldsymbol{\sigma} = \lambda \mathbf{F} \mathbf{F}^T$ .
- (b)  $\boldsymbol{\sigma} = \lambda [\nabla \mathbf{u} + (\nabla \mathbf{u})^T]$ .
- (c)  $\boldsymbol{\sigma} = \lambda I_B \mathbf{I}$ , where  $\mathbf{B} = \mathbf{F} \mathbf{F}^T$  and  $I_B = \text{tr}(\mathbf{B})$ .
- (d)  $\boldsymbol{\sigma} = \lambda (\mathbf{I} - \mathbf{B}^{-1})$ , where  $\mathbf{B} = \mathbf{F} \mathbf{F}^T$ .

**8.15** This problem derives formulas for vector and tensor quantities when making the rigid body change of variables given in (8.50).

- (a) Show that  $\mathbf{F}^* = \mathbf{Q} \mathbf{F} \mathbf{Q}_0^T$ .
- (b) Show that  $\mathbf{U}^* = \mathbf{Q} \mathbf{U} + (\mathbf{Q} - \mathbf{Q}_0) \mathbf{X} + \mathbf{b} - \mathbf{b}_0$ .
- (c) Show that  $(\nabla_X \mathbf{U})^* = \mathbf{Q} (\nabla_X \mathbf{U}) \mathbf{Q}_0^T + \mathbf{Q} \mathbf{Q}_0^T - \mathbf{I}$ .
- (d) Show that  $(\nabla \mathbf{u})^* = \mathbf{Q}_0 (\nabla \mathbf{u}) \mathbf{Q}^T + \mathbf{I} - \mathbf{Q}_0 \mathbf{Q}^T$ .
- (e) Show that  $\mathbf{D}^* = \mathbf{Q} \mathbf{D} \mathbf{Q}^T$  and  $\mathbf{W}^* = \mathbf{Q} \mathbf{W} \mathbf{Q}^T + \boldsymbol{\Omega}$ .
- (f) Show that  $(\nabla \cdot \boldsymbol{\sigma})^* = \mathbf{Q} (\nabla \cdot \boldsymbol{\sigma})$ .
- (g) Setting  $\mathbf{c} = \mathbf{Q}^T \mathbf{b}$ , show that  $\mathbf{V} = \mathbf{Q}^T \mathbf{V}^* + (\mathbf{Q}^T)' \mathbf{x}^* - \mathbf{c}'$ . From this show that

$$\frac{D\mathbf{v}}{Dt} = \mathbf{Q}^T \left( \frac{D\mathbf{v}^*}{Dt} \right)^* + 2(\mathbf{Q}^T)' \mathbf{v}^* + (\mathbf{Q}^T)'' \mathbf{x}^* - \mathbf{c}''.$$

- (h) Assuming the body force transforms as  $\mathbf{f}^* = \mathbf{Q}\mathbf{f}$ , and the density as  $\rho^* = \rho$ , show that the momentum equation in the  $\mathbf{x}^*$  coordinate system is

$$\rho^* \left( \frac{D\mathbf{v}}{Dt} \right)^* = (\nabla \cdot \boldsymbol{\sigma})^* + \rho^* \mathbf{f}^* - \rho^* \mathbf{z}^*,$$

where  $\mathbf{z}^* = 2\mathbf{Q}(\mathbf{Q}^T)' \mathbf{v}^* + \mathbf{Q}\mathbf{Q}'' \mathbf{x}^* - \mathbf{Q}\mathbf{c}''$  is an acceleration term that comes from the change of variables.

- (i) Explain why the momentum equation is not Euclidian invariant but is Galilean invariant.

**8.16** This problem develops some of the properties of the principal invariants of a symmetric matrix  $\mathbf{R}$ .

- (a) Derive the formulas in (8.62)–(8.64).  
 (b) Show that the characteristic equation for  $\mathbf{R}$  can be written as  $\lambda^3 - \text{I}_R \lambda^2 + \text{II}_R \lambda - \text{III}_R = 0$ .  
 (c) By definition,  $\text{I}_R$  is a function of the components of  $\mathbf{R}$ . Show that

$$\frac{\partial}{\partial R_{ij}} \text{I}_R = \delta_{ij},$$

where  $\delta_{ij}$  is the Kronecker delta function. Explain why the above equation can be written as

$$\frac{\partial}{\partial \mathbf{R}} \text{I}_R = \mathbf{I}.$$

- (d) Using the ideas developed in part (c) show that

$$\frac{\partial}{\partial \mathbf{R}} \text{II}_R = -\mathbf{R} + \text{I}_R \mathbf{I}.$$

## Section 8.10

**8.17** This problem considers the simple shear motion of an incompressible fluid as given in Sect. 8.2.

- (a) What is the stress?  
 (b) What is the corresponding spatial velocity  $\mathbf{v}$ ? Assuming there are no external forces, when will this be a solution of the equations of motion?

**8.18** This problem considers the rigid body motion of an incompressible fluid as given in Sect. 8.2.

- (a) What is the stress?

- (b) What is the corresponding spatial velocity  $\mathbf{v}$ ? When will this be a solution of the equations of motion?

**8.19** This problem considers fluid flow that is also a linear flow (this is defined in Exercise 8.5).

- Show that linear flow is possible for an incompressible material only if  $\text{tr}(\mathbf{H}) = 0$ .
- Assuming there are no body forces, show that linear flow is a solution of the incompressible Navier-Stokes equations if  $\mathbf{H}'(t)$  is symmetric and  $\text{tr}(\mathbf{H}) = 0$ .
- Under what conditions, if any, does simple shear satisfy the conditions in part (b)?
- Under what conditions, if any, does rigid body motion satisfy the conditions in part (b)?

**8.20** Suppose that the incompressible fluid equations (8.77) and (8.78) are to be solved in a bounded domain  $D$ , and the impermeability condition (8.79) is used on the boundary  $\partial D$ . Show that not just any boundary velocity  $\mathbf{v}_s$  can be used. Namely, show that the given velocity must satisfy

$$\iint_{\partial D} \mathbf{v}_s \cdot \mathbf{n} dS = 0.$$

What is the physical meaning of the above equation?

## Section 8.12

**8.21** Show that in material coordinates the assumption of incompressibility results in the equation  $\det(\mathbf{F}) = 1$ .

**8.22** This problem considers the simple shear motion given in Sect. 8.2.

- Find the Green strain tensor.
- What is the first Piola-Kirchhoff stress tensor, as given in (8.99)?
- What is  $\mathbf{U}$ ? Assuming there are no external forces, when will this be a solution of the equations of motion?
- Find the polar decomposition  $\mathbf{F} = \mathbf{Q}\mathbf{R}$ . Note that your answer will depend on whether  $\alpha$  is positive or negative.

**8.23** This problem considers rigid body motion as given in Sect. 8.2.

- What is the first Piola-Kirchhoff stress tensor, as given in (8.99)?
- What is  $\mathbf{U}$ ? Assuming there are no external forces, when will this be a solution of the equations of motion?

**8.24** Determine if the following can be used for constitutive laws. This means that (8.94) and (8.95) are satisfied.

- (a)  $\mathbf{P} = \alpha(\mathbf{F} + \mathbf{F}^T)$ .
- (b)  $\mathbf{P} = \alpha \det(\mathbf{F})(\mathbf{F} - \mathbf{F}^{-T})$ , where  $\mathbf{F}^{-T} = (\mathbf{F}^T)^{-1} = (\mathbf{F}^{-1})^T$ .
- (c)  $\boldsymbol{\sigma} = \alpha I_A \mathbf{I}$ , where  $\mathbf{A} = \frac{1}{2}(\mathbf{I} - \mathbf{B}^{-1})$  and  $\mathbf{B} = \mathbf{F}\mathbf{F}^T$ .

**8.25** Suppose that an elastic bar is stretched, or compressed, from length  $\ell_0$  to  $\ell$  (see Table 6.3). This is referred to as simple extension, and it corresponds to having  $x = \alpha X$ ,  $y = \beta Y$ , and  $z = \beta Z$ , where  $\alpha = \ell/\ell_0$ . The value for  $\beta$  depends on the constitutive law used for the stress. It is required that  $\beta > 0$ , and, in theory,  $0 < \alpha < \infty$ .

- (a) Determine  $\mathbf{F}$ ,  $\mathbf{F}^T \mathbf{F}$ , and  $\mathbf{F}\mathbf{F}^T$ .
- (b) Using the constitutive law (8.99), and your results from part (a), find  $\mathbf{P}$ .
- (c) It is assumed that there are no forces applied to the bar along its lateral surface. This means that it is required that  $P_{22} = P_{33} = 0$ . Use this to find  $\beta$ .
- (d) Using your results from parts (b) and (c), show that

$$P_{11} = \frac{\mu_s(3\lambda_s + 2\mu_s)}{2(\mu_s + \lambda_s)} \alpha(\alpha^2 - 1).$$

Explain why this requires that  $\lambda_s \alpha^2 \leq 2\mu_s + 3\lambda_s$ .

- (e) A possible constitutive law is  $\boldsymbol{\sigma} = \lambda_s I_A \mathbf{I} + 2\mu_s \mathbf{A}$ , where  $\mathbf{A} = \frac{1}{2}(\mathbf{I} - \mathbf{B}^{-1})$  is the Almansi strain tensor and  $\mathbf{B} = \mathbf{F}\mathbf{F}^T$ . Use your results from part (a) to find  $\mathbf{P}$ , and then redo part (c) to find  $\beta$ . With this, show that

$$P_{11} = \mu_s \left( 1 - \frac{2(\lambda_s + \mu_s)}{(2\mu_s + 3\lambda_s)\alpha^2 - \lambda_s} \right).$$

Explain why this requires that  $\lambda_s < (2\mu_s + 3\lambda_s)\alpha^2$ .

- (f) A possible constitutive law is  $\mathbf{S} = \lambda_s I_H \mathbf{I} + 2\mu_s \mathbf{H}$ , where  $\mathbf{H} = \frac{1}{2} \ln(\mathbf{F}^T \mathbf{F})$  is the Hencky strain tensor. The definition of the logarithm of a matrix is given in Exercise 8.27(c). Use your results from part (a) to find  $\mathbf{P}$ , and then redo part (c) to find  $\beta$ . With this, show that

$$P_{11} = \frac{\mu_s(3\lambda_s + 2\mu_s)}{(\mu_s + \lambda_s)} \alpha \ln \alpha.$$

Hint: If  $\mathbf{A}$  is diagonal, then  $\ln(\mathbf{A})$  is diagonal.

- (g) On the same axes, sketch the  $P_{11}$ 's obtained in parts (d), (e), and (f) as a function of  $\alpha$ . Based on what you expect happens with an elastic bar, which constitutive law is the better choice? Make sure to explain why.

**8.26** A strain tensor  $\mathbf{Z}$  must be zero if there is rigid body motion. For rigid body motion,  $\mathbf{X} = \bar{\mathbf{Q}}\mathbf{X} + \mathbf{b}$ , where  $\bar{\mathbf{Q}}$  is a rotation matrix. Also, according to the polar decomposition theorem,  $\mathbf{F} = \mathbf{Q}\mathbf{R}$  and  $\mathbf{F} = \mathbf{L}\mathbf{Q}$ , where  $\mathbf{Q}$  is a rotation matrix, and  $\mathbf{R}$

and  $\mathbf{L}$  are symmetric and positive definite. It is helpful to know that  $\mathbf{Q}$ ,  $\mathbf{R}$ , and  $\mathbf{L}$  are unique.

- What is  $\mathbf{R}$ , and what is  $\mathbf{L}$ , for a rigid body motion?
- Show that the Green strain tensor can be written as  $\mathbf{E} = \frac{1}{2}(\mathbf{R}^2 - \mathbf{I})$ . Use this to show that it satisfies the stated conditions for a strain.
- Show that  $\mathbf{Z} = \nabla_{\mathbf{X}}\mathbf{U}$  does not qualify as a strain tensor.
- Show that  $\mathbf{Z} = \mathbf{B}$ , where  $\mathbf{B} = \mathbf{F}\mathbf{F}^T$ , does not qualify as a strain tensor.
- Show that  $\mathbf{Z} = \frac{1}{2}(\mathbf{I} - \mathbf{B}^{-1})$  satisfies the stated conditions. This is known as the Almansi strain tensor.

**8.27** If  $\mathbf{S}$  is assumed to be a linear function of a given strain tensor  $\mathbf{Z}$ , then the constitutive law is  $\mathbf{S} = \lambda I_Z \mathbf{I} + 2\mu \mathbf{Z}$ , where  $I_Z = \text{tr}(\mathbf{Z})$ , and  $\lambda$  and  $\mu$  are constants. Assuming that  $\mathbf{Z}$  is symmetric, then  $\mathbf{S}$  is symmetric. So, to satisfy frame-invariance, it is required that  $\mathbf{S}^* = \mathbf{Q}_0 \mathbf{S} \mathbf{Q}_0^T$ .

- Show that if  $\mathbf{Z}^* = \mathbf{Q}_0 \mathbf{Z} \mathbf{Q}_0^T$ , then  $\mathbf{S}^* = \mathbf{Q}_0 \mathbf{S} \mathbf{Q}_0^T$ .
- Show that  $\mathbf{E}^* = \mathbf{Q}_0 \mathbf{E} \mathbf{Q}_0^T$ .
- The Hencky strain tensor is  $\mathbf{H} = \frac{1}{2} \ln(\mathbf{F}^T \mathbf{F})$ . The logarithm involves the Taylor series  $\ln(x) = (x - 1) - \frac{1}{2}(x - 1)^2 + \frac{1}{3}(x - 1)^3 - \frac{1}{4}(x - 1)^4 + \dots$ . So, given a square matrix  $\mathbf{A}$ ,

$$\ln(\mathbf{A}) \equiv (\mathbf{A} - \mathbf{I}) - \frac{1}{2}(\mathbf{A} - \mathbf{I})^2 + \frac{1}{3}(\mathbf{A} - \mathbf{I})^3 - \frac{1}{4}(\mathbf{A} - \mathbf{I})^4 + \dots$$

Show that, given an orthogonal matrix  $\mathbf{Q}$ ,  $\ln(\mathbf{Q}\mathbf{A}\mathbf{Q}^T) = \mathbf{Q} \ln(\mathbf{A}) \mathbf{Q}^T$ . With this, show that  $\mathbf{H}^* = \mathbf{Q}_0 \mathbf{H} \mathbf{Q}_0^T$ .

## Section 8.13

**8.28** The kinetic energy for a regular region  $R(t)$  is

$$K = \iiint_{R(t)} \frac{1}{2} \rho \mathbf{v} \cdot \mathbf{v} dV.$$

Let  $K_0$  be the value of  $K$  when the motion is irrotational, which means that the velocity can be written as  $\mathbf{v} = \nabla \phi$ . Let  $\mathbf{v}$  be any other velocity, not necessarily irrotational, but which has the same normal velocity at the boundary as the irrotational motion. This means that  $\mathbf{v} \cdot \mathbf{n} = (\nabla \phi) \cdot \mathbf{n}$  on  $\partial R$ . Assuming the density is constant, show that  $K_0 \leq K$ . This observation that irrotational flows minimize the kinetic energy is known as Kelvin's Minimum Energy Theorem.



**8.29** This problem extends some of the ideas developed with the energy equation.

- Using the second law of thermodynamics it can be shown that the stress power is nonnegative. Use this to show that if  $\boldsymbol{\sigma} = -p\mathbf{I} + \alpha_0\mathbf{I} + \alpha_1\mathbf{D} + \alpha_2\mathbf{D}^2$ , and the material is incompressible, then  $0 \leq \alpha_1 \text{tr}(\mathbf{D}^2) + 3\alpha_2 \det(\mathbf{D})$ . Hint: The Cayley-Hamilton theorem will be useful here.
- Use the result from part (a) to prove that the dynamic viscosity of an incompressible Newtonian fluid is nonnegative.
- Show that the potential energy density (8.118) can be written as

$$U_p = \frac{1}{8}(\lambda_s + 2\mu_s)\mathbf{I}_B^2 - \frac{1}{4}(3\lambda_s + 2\mu_s)\mathbf{I}_B - \frac{1}{2}\mu_s\Pi_B.$$

**8.30** This problem provides some of the formulas used in the derivation of the energy equation.

- Assuming a vector  $\mathbf{v}$  and tensor  $\boldsymbol{\sigma}$  are functions of  $\mathbf{x}$ , show that

$$\mathbf{v} \cdot (\nabla \cdot \boldsymbol{\sigma}) = \nabla \cdot (\boldsymbol{\sigma} \mathbf{v}) - \text{tr}(\boldsymbol{\sigma} \nabla \mathbf{v}).$$

Also, explain why this holds even if  $\boldsymbol{\sigma}$  is not symmetric.

- If  $\boldsymbol{\sigma}$  is symmetric show that  $\text{tr}(\boldsymbol{\sigma} \nabla \mathbf{v}) = \text{tr}(\boldsymbol{\sigma} (\nabla \mathbf{v})^T)$ . With this, show that  $\mathbf{v} \cdot (\nabla \cdot \boldsymbol{\sigma}) = \nabla \cdot (\boldsymbol{\sigma} \mathbf{v}) - \text{tr}(\boldsymbol{\sigma} \mathbf{D})$ , where  $\mathbf{D} = \frac{1}{2}(\nabla \mathbf{v} + (\nabla \mathbf{v})^T)$ .
- Explain why part (a) can be used to show that  $\mathbf{V} \cdot (\nabla_X \cdot \mathbf{P}^T) = \nabla_X \cdot (\mathbf{P}^T \mathbf{V}) - \text{tr}(\mathbf{P}^T \nabla_X \mathbf{V})$ .

### Additional Questions

**8.31** This problem concerns the inverse of the constitutive law for a linear material.

- For a viscous compressible fluid show that

$$\mathbf{D} = -\frac{\lambda}{2\mu(3\lambda + 2\mu)}(3p + \text{tr}(\boldsymbol{\sigma}))\mathbf{I} + \frac{1}{2\mu}(\boldsymbol{\sigma} + p\mathbf{I}).$$

- For a linearly elastic material show that

$$\mathbf{E} = -\frac{\lambda_s}{2\mu_s(3\lambda_s + 2\mu_s)}\text{tr}(\mathbf{P})\mathbf{I} + \frac{1}{2\mu_s}\mathbf{P}.$$

**8.32** This problem derives the Rivlin-Ericksen representation theorem for a special case. The assumption is that the constitutive law for the stress is  $\boldsymbol{\sigma} = \mathbf{G}(\mathbf{R})$ , where  $\mathbf{R}$  is a symmetric tensor. Also, assume that the function  $\mathbf{G}$  can be expanded using a Taylor series to give

$$\mathbf{G} = \sum_{n=0}^{\infty} \kappa_n \mathbf{R}^n,$$

where the  $\kappa_n$ 's are constants and  $\mathbf{R}^n = \mathbf{I}$  for  $n = 0$ .

- (a) Use the Cayley-Hamilton theorem to show that the constitutive law can be written as  $\boldsymbol{\sigma} = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{R} + \alpha_2 \mathbf{R}^2$ , where the coefficients  $\alpha_0$ ,  $\alpha_1$ , and  $\alpha_2$  are functions of the three invariants of  $\mathbf{R}$ .
- (b) Suppose  $\mathbf{R}$  is objective, so that  $\mathbf{R}^* = \mathbf{Q}\mathbf{R}\mathbf{Q}^T$ . Does the assumed constitutive law satisfy the Principle of Material Frame-Indifference?

## Chapter 9

# Newtonian Fluids



The equations of motion for an incompressible Newtonian fluid are given in Sect. 8.11.1. They were derived based on the assumption that the constitutive law for the stress is

$$\boldsymbol{\sigma} = -p\mathbf{I} + 2\mu\mathbf{D}, \quad (9.1)$$

where  $p$  is the pressure,  $\mathbf{D}$  is the rate of deformation tensor given in (8.66), and  $\mu$  is the dynamic viscosity. The SI unit for  $\mu$  is the Pascal-second (Pa·s), and to help provide some perspective on this, the viscosities of some well-known fluids are given in Table 9.1. Not unexpectedly, the viscosity of air is significantly less than the viscosity of water, which in turn is less viscous than olive oil and honey.

The fluids listed in Table 9.1 are Newtonian, but many fluids are not. Examples of non-Newtonian fluids include ketchup, yogurt, toothpaste, and peanut butter. This raises the important question of how to determine if a fluid can be modeled as a Newtonian fluid. A similar question came up in Chaps. 6 and 7 when studying elasticity and viscoelasticity, and the answer is the same as before. Namely, we will derive solutions to the equations of motion and then compare them with what is found experimentally. Assuming they agree, then we should be able to use the experimental data to determine the viscosity.

The complication is that it is not possible to solve the equations for a Newtonian fluid, by hand, without making simplifying assumptions about what problem is going to be solved. As an example, the viscosity of air is so small, it would seem that it might be possible to simply assume it is zero. This assumption produces what is known as an inviscid fluid, and the resulting mathematical problem gives the appearance of being simpler than what is obtained for a viscous fluid.

In this chapter various simplifying assumptions are examined, with the goal of better understanding fluid motion. The first concerns what is known as steady flow, which means the velocity is independent of time. This assumption, by itself, does not yield easily solved problems. So, we will restrict the motion to be mainly

**Table 9.1** Viscosity and density of various fluids at 25 °C

Fluid	Viscosity (Pa·s)	Density (kg/m <sup>3</sup> )
Air	$1.8 \times 10^{-5}$	1.18
Water	$8.9 \times 10^{-4}$	$0.997 \times 10^3$
Mercury	$1.5 \times 10^{-3}$	$1.3 \times 10^4$
Olive oil	0.08	$0.92 \times 10^3$
Honey	37	$1.4 \times 10^3$

unidirectional, which means flow in a straight pipe or along a flat plate. This covers most of the examples considered in Sects. 9.1 and 9.5. To explain the second simplification, fluid flow can be rotational, directional, or, more often, a combination of both. We will consider the extreme cases, when it is just rotational, and when it is just directional. These are the topics considered in Sects. 9.2 and 9.3. The third simplification concerns the viscosity. We will consider, in Sect. 9.4, what happens when the viscosity is zero, and, in Sect. 9.5, how to solve the problem when the viscosity is nonzero but relatively small.

## 9.1 Steady Flow

One of the more studied problems in fluids involves steady flow. This means that the fluid velocity and pressure are independent of time. Assuming there are no body forces, then the equations of motion for a steady incompressible fluid, coming from (8.77) and (8.78), are

$$\rho(\mathbf{v} \cdot \nabla)\mathbf{v} = -\nabla p + \mu \nabla^2 \mathbf{v}, \quad (9.2)$$

$$\nabla \cdot \mathbf{v} = 0. \quad (9.3)$$

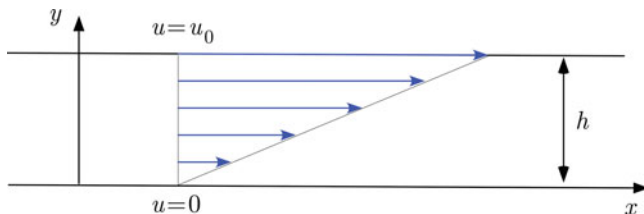
As always, with incompressible motion, it is assumed that  $\rho$  is constant.

### 9.1.1 Plane Couette Flow

One of the more basic flows arises when studying the motion of a fluid between two parallel plates. A cross-section of this configuration is shown in Fig. 9.1. The lower plate, located at  $y = 0$ , is fixed, while the upper plate, at  $y = h$ , moves with a constant velocity  $u_0$  in the  $x$ -direction. From the no-slip condition, the boundary conditions are

$$\mathbf{v} = (u_0, 0, 0) \quad \text{on } y = h, \quad (9.4)$$

$$\mathbf{v} = \mathbf{0} \quad \text{on } y = 0. \quad (9.5)$$



**Fig. 9.1** In plane Couette flow the lower plate is stationary, while the top plate moves in the  $x$ -direction. Solving this problem shows that the velocity of the fluid varies linearly between the two plates

It is assumed that the upper plate has been moving with this constant velocity for a long time, so the flow is steady. It is also assumed that the fluid is incompressible, so (9.2) and (9.3) apply. The resulting problem corresponds to what is called *plane Couette flow*.

At first glance, given that (9.2) is a nonlinear partial differential equation, finding the velocity and pressure would seem to be an almost impossible task. However, some useful insights on the properties of the solution can be derived from the boundary conditions and the geometry. In particular, given that the upper and lower boundaries are flat plates, and the upper one moves with a constant velocity in the  $x$ -direction, it is not unreasonable to guess that there is no dependence on, or motion in, the  $z$ -direction. In other words  $\mathbf{v} = (u, v, 0)$ , where  $u$ ,  $v$ , and  $p$  are independent of  $z$ . In this case, (9.2) and (9.3) reduce to

$$\begin{aligned}\rho(u\partial_x + v\partial_y)u &= -\partial_x p + \mu(\partial_x^2 + \partial_y^2)u, \\ \rho(u\partial_x + v\partial_y)v &= -\partial_y p + \mu(\partial_x^2 + \partial_y^2)v, \\ \partial_x u + \partial_y v &= 0.\end{aligned}$$

This is still a formidable problem, so we need another insight into the form of the solution. Given that the upper plate is sliding in the  $x$ -direction, it is not unreasonable to expect that there is no flow in the  $y$ -direction. In this case,  $v = 0$  and the above system reduces to

$$\begin{aligned}\rho u \partial_x u &= -\partial_x p + \mu(\partial_x^2 + \partial_y^2)u, \\ 0 &= -\partial_y p, \\ \partial_x u &= 0.\end{aligned}$$

From the last two equations we have that  $p = p(x)$  and  $u = u(y)$ . In this case, the first equation reduces to  $p'(x) = \mu u''(y)$ . The only way for a function of  $x$  to equal a function of  $y$  is that both functions are constants. Consequently,  $p'(x)$  constant means  $p(x) = p_0 + xp_1$ , where  $p_0$  and  $p_1$  are constants. It is assumed that the

pressure remains bounded, and so  $p_1 = 0$ . With this, the solution of  $\mu u''(y) = p'(x)$  is  $u = ay + b$ . Imposing the boundary conditions  $u(0) = 0$  and  $u(h) = u_0$ , it follows that  $u = u_0 y/h$ .

The solution of the plane Couette flow problem is, therefore,

$$\mathbf{v} = (\dot{\gamma}y, 0, 0), \quad (9.6)$$

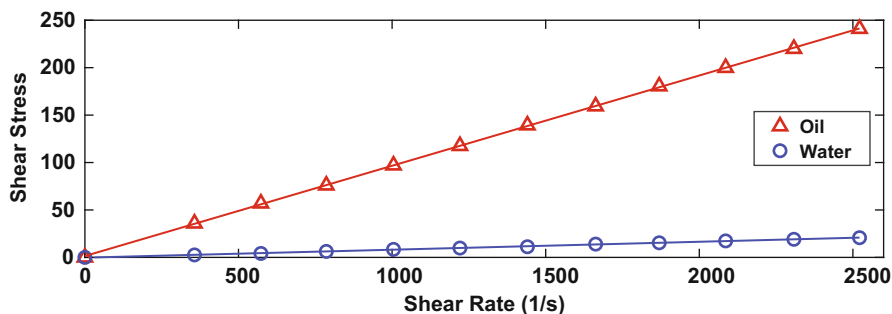
where  $p = p_0$  is constant, and

$$\dot{\gamma} = \frac{u_0}{h} \quad (9.7)$$

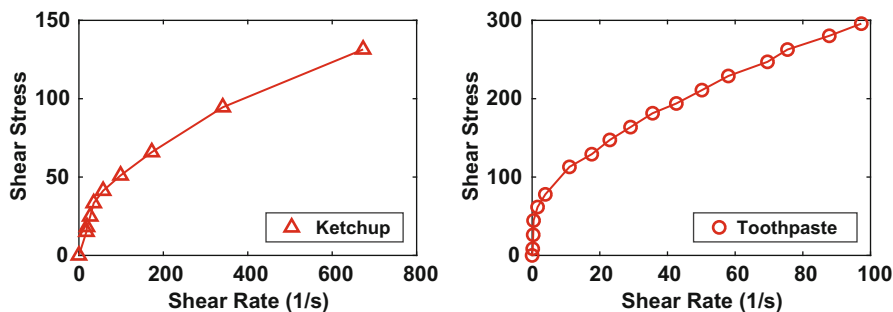
is known as the *shear rate*. This shows that the fluid velocity in the  $x$ -direction increases linearly between the two plates, from zero to  $u_0$ . This dependence is illustrated in Fig. 9.2. Also, the resulting fluid stress tensor (9.1) is

$$\boldsymbol{\sigma} = -p_0 \mathbf{I} + \mu \begin{pmatrix} 0 & \frac{u_0}{h} & 0 \\ \frac{u_0}{h} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (9.8)$$

The above solution gives us something we sorely need, and that is a method for checking on the assumption that a fluid is Newtonian. The solution shows that for a Newtonian fluid the shear stress is  $\sigma_{12} = \mu u_0/h$ . Therefore, the shear stress depends linearly on the shear rate  $\dot{\gamma} = u_0/h$ , with the slope of the line equal to the viscosity. This is the basis for one of the more important experiments in fluid dynamics, where the shear stress is measured as a function of the shear rate. Results from such tests are shown in Figs. 9.2 and 9.3, for fluids most people have experience with. Based on the linearity of the data in Fig. 9.2, the assumption that water and oil are Newtonian is reasonable. For the same reason, from Fig. 9.3, ketchup and toothpaste are not



**Fig. 9.2** Shear stress, as a function of shear rate  $\dot{\gamma}$ , for water and oil at 25 °C (Ellenberger et al. 1976)



**Fig. 9.3** Shear stress, as a function of shear rate  $\dot{\gamma}$ , for ketchup and toothpaste at room temperature (Leong and Yeow 2003)

Newtonian, or only Newtonian for very small shear rates. They are examples of what are called nonlinear power-law fluids, where  $\sigma_{12} = \alpha \dot{\gamma}^\beta$ . Based on the data in Fig. 9.3, for ketchup,  $\beta = 0.55$ , whereas for toothpaste,  $\beta = 0.44$ . Some of the implications of such a constitutive law are investigated in Exercise 9.25.

Before moving on to another topic, a comment needs to be made about our solution of the plane Couette flow problem. The assumptions we made in deriving the solutions worked in the sense that we found pressure and velocity functions that satisfy the original steady flow problem given in (9.2) and (9.3), along with the stated boundary conditions (9.4) and (9.5). So, if the solution to this problem is unique, then we have found it. The question is uniqueness. We saw in Chap. 3 that nonlinear problems often have multiple solutions. In such cases the question that arose was whether the solution was asymptotically stable, because even if there are multiple solutions, those that are unstable are effectively unachievable. This question arises in all but the simplest fluid problems because of the inherent nonlinear nature of fluid flow. It has been shown that the solution we have derived for plane Couette flow is linearly stable (Drazin and Reid 2004). However, it has been found experimentally that as the shear rate increases there is a value where the flow changes dramatically, from the unidirectional flow we found to one that is three-dimensional and turbulent. The appearance of a solution different than the one we derived is due to the experimental setup. It is not possible to have infinite flat plates, or strictly two-dimensional flow, and the effects of these perturbations become more pronounced at larger shear rates. A discussion of the difficulty of doing plane Couette flow experiments, and the appearance of a turbulent flow at higher shear rates, can be found in Tillmark and Alfredsson (1992).

### 9.1.2 Poiseuille Flow

A second steady flow that is often studied involves the fluid motion through a long pipe due to a pressure difference between the two ends. The pipe is assumed to have length  $L$ , and radius  $R$ . Given the geometry, it is easier to use cylindrical coordinates, with the pipe oriented as shown in Fig. 9.4. In this case the spatial coordinate system is  $(r, \theta, z)$ , and the associated velocity vector is  $\mathbf{v} = (v_r, v_\theta, v_z)$ . It should be noted that the subscripts on the components of this vector do not indicate differentiation, but identify the coordinate of the particular velocity component. So, for example,  $v_r$  is the velocity in the  $r$ -direction.

The boundary conditions for Poiseuille flow are

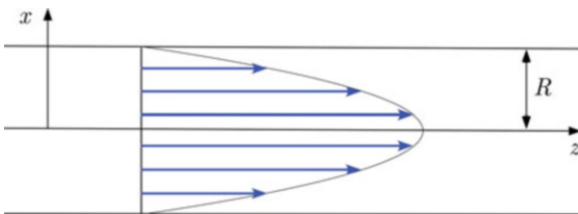
$$\mathbf{v} = \mathbf{0} \quad \text{at } r = R, \quad (9.9)$$

$$p = p_0, \quad v_r = v_\theta = 0 \quad \text{at } z = 0, \quad (9.10)$$

$$p = p_1, \quad v_r = v_\theta = 0 \quad \text{at } z = L. \quad (9.11)$$

To explain these, (9.9) is the no-slip condition and it applies because the pipe is not moving. The conditions at  $z = 0$  and  $z = L$  account for the prescribed constant pressures at these ends, and the assumption that the fluid velocity is only in the axial direction as it enters and leaves the pipe. The resulting problem corresponds to what is called *Poiseuille flow*.

To find the solution, we will first consider some of the basic properties of the flow. Given the boundary conditions (9.9) and (9.10), both  $v_r$  and  $v_\theta$  are zero on the pipe and at both ends. Based on this, it is expected that  $v_r = v_\theta = 0$  everywhere. Also, there is no  $\theta$  dependence in the boundary, or the boundary conditions. Because of this it is expected that the axial velocity  $v_z$  and pressure  $p$  do not depend on the angular coordinate  $\theta$ . In other words, it is expected that  $v_z = v_z(r, z)$  and  $p = p(r, z)$ . The equations of motion in cylindrical coordinates, which are given in Appendix E, in this case reduce to



**Fig. 9.4** In Poiseuille flow, fluid moves through a pipe due to a pressure difference across the ends. Solving this problem shows that the axial velocity of the fluid has a parabolic distribution, as given in (9.13)



$$\begin{aligned}
0 &= \frac{\partial p}{\partial r}, \\
\rho v_z \frac{\partial v_z}{\partial z} &= -\frac{\partial p}{\partial z} + \mu \left( \frac{\partial^2 v_z}{\partial r^2} + \frac{1}{r} \frac{\partial v_z}{\partial r} + \frac{\partial^2 v_z}{\partial z^2} \right), \\
\frac{\partial v_z}{\partial z} &= 0.
\end{aligned} \tag{9.12}$$

From the first and third equation we conclude that  $p = p(z)$  and  $v_z = v_z(r)$ . In this case (9.12) reduces to

$$\frac{dp}{dz} = \mu \left( \frac{d^2 v_z}{dr^2} + \frac{1}{r} \frac{dv_z}{dr} \right).$$

The left-hand side is only a function of  $z$ , while the right-hand side is only a function of  $r$ . The only way that this can happen is that  $p'(z)$  is constant. Given the boundary conditions on the pressure we conclude that  $p = p_0 + z(p_1 - p_0)/L$ . The remaining equation (9.12) reduces to

$$\mu \left( \frac{d^2 v_z}{dr^2} + \frac{1}{r} \frac{dv_z}{dr} \right) = p_1/L.$$

This is a first order equation for  $\frac{d}{dr}v_z$ . Using this observation to solve the equation, one finds that the general solution is

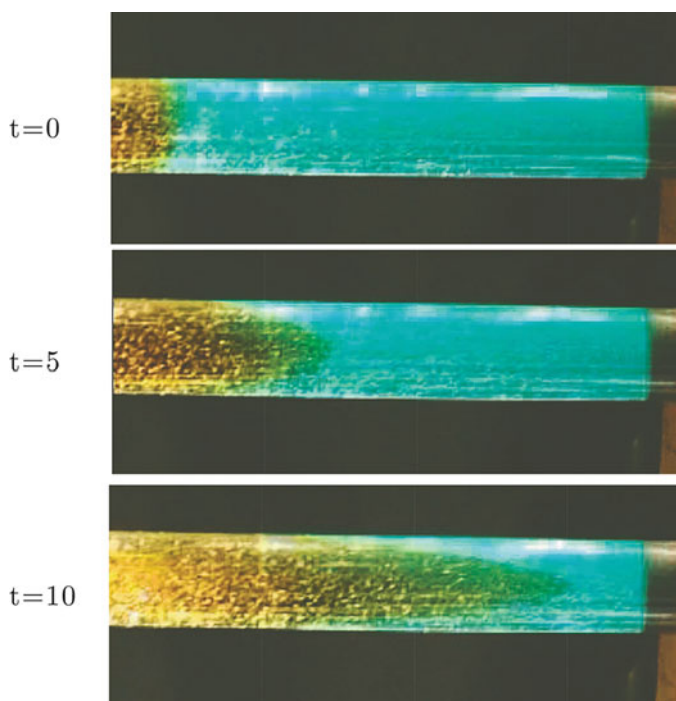
$$v_z = \frac{p_1 - p_0}{4\mu L} r^2 + a \ln(r) + b.$$

The solution must be bounded, so  $a = 0$ , and it must also satisfy the no-slip boundary condition  $v_z = 0$  at  $r = R$ . The resulting axial velocity is therefore

$$v_z = \frac{p_0 - p_1}{4\mu L} (R^2 - r^2). \tag{9.13}$$

This shows that the velocity has a parabolic distribution in the pipe, and this is illustrated in Fig. 9.4. The fact that pipe flow has this parabolic shape is demonstrated in Fig. 9.5.

It is important to make a point that was also made after solving the plane Couette flow problem. Several simplifying assumptions were made about the velocity and pressure functions, based on the given boundary conditions and geometry of the pipe, to reduce the momentum equations down to (9.12). These assumptions might be better described as educated guesses of the form of the solution. They worked in the sense that we found pressure and velocity functions that satisfy the original



**Fig. 9.5** Two fluids flowing, from left to right, in a clear pipe (Kunkle 2008). At  $t = 0$  the darker fluid is located at the left end. At  $t = 10$  s the darker fluid shows the parabolic shape predicted by the solution given in (9.13)

steady flow problem given in (9.2) and (9.3), along with the stated boundary conditions (9.9)–(9.11). So, if the solution to this problem is unique, then we have found it. Moreover, an experimental demonstration that the solution has the predicted parabolic profile is shown in Fig. 9.5.

As it turns out, experiments show that non-parabolic flow can be obtained in pipe flow. As with the plane Couette problem, for large enough perturbations in the flow, it is found that at high velocities the flow in the pipe can be three-dimensional and turbulent. However, based on numerical studies and very careful experiments, it is widely accepted that the parabolic solution is linearly stable irrespective of the velocity. How turbulence arises in the flow is not understood, and this is considered to be one of the major open problems in hydrodynamic stability theory. A review of what is known, and some of the theoretical and experimental questions that need to be answered, can be found in Eckhardt et al. (2007), Willis et al. (2008), and Mullin (2011).

## 9.2 Vorticity

If you float on an inner tube on a river you notice that not only do you move downstream, the moving water also causes you to spin. It is the rotational component of the motion that we are now interested in exploring. The first step is to derive a variable that can be used to measure the rotation, at least locally.

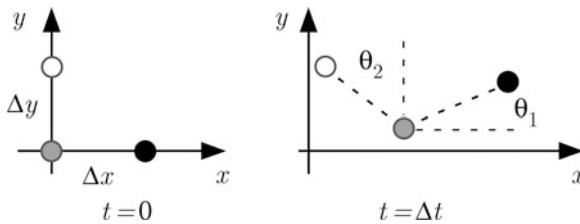
To explain how this is done, consider three fluid particles located on the coordinate axes at  $t = 0$ , and a short time later, at  $t = \Delta t$ , as shown in Fig. 9.6. For simplicity the flow is assumed to be two-dimensional, so the fluid velocity is  $\mathbf{v} = (u(x, y), v(x, y))$ . At  $t = 0$ , the velocity of the particle located at the origin is  $\mathbf{v}_0 = (u_0, v_0)$ , where  $u_0 = u(0, 0)$  and  $v_0 = v(0, 0)$ . At  $t = \Delta t$ , its position is, approximately,  $\Delta t \mathbf{v}_0$ . For the particle that is initially located on the  $x$ -axis, at  $x = \Delta x$ , its velocity at  $t = 0$  is, using Taylor's theorem,

$$\mathbf{v}_1 = (u(\Delta x, 0), v(\Delta x, 0)) \approx \mathbf{v}_0 + \Delta x(u_x(0, 0), v_x(0, 0)).$$

At  $t = \Delta t$ , its position is, approximately,  $\Delta t \mathbf{v}_1$ . With this

$$\begin{aligned} \tan(\theta_1) &\approx \frac{\Delta t(v_0 + \Delta x v_x) - \Delta t v_0}{\Delta x + \Delta t(u_0 + \Delta x u_x) - \Delta t u_0} \\ &= \frac{\Delta t v_x}{1 + \Delta t u_x} \\ &\approx \Delta t v_x. \end{aligned}$$

Given that the angle is small, so  $\tan(\theta_1) \approx \theta_1$ , we have that  $\theta_1 \approx \Delta t v_x$ . Carrying out a similar analysis using the particle that started at  $y = \Delta y$  one finds that  $\theta_2 \approx -\Delta t u_y$ . The average angular velocity around the  $z$ -axis is therefore  $(\theta_1 + \theta_2)/(2\Delta t) \approx \frac{1}{2}(v_x - u_y)$ . Similar expressions can be derived for the rotation around the other two axes.



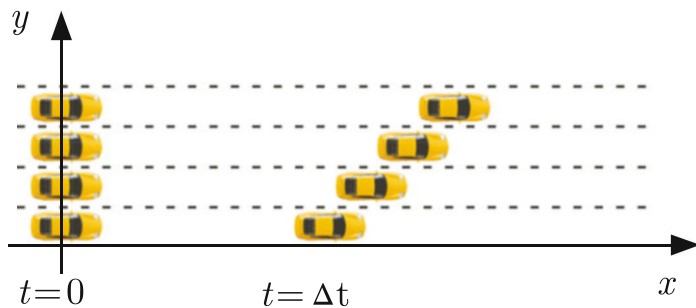
**Fig. 9.6** Three nearby fluid particles used to introduce the concept of vorticity. Their motion from  $t = 0$  to  $t = \Delta t$ , causes both a translation and relative rotation in their configuration

This is the motivation for introducing the *vorticity* vector  $\boldsymbol{\omega}$ , which is defined as

$$\begin{aligned}\boldsymbol{\omega} &\equiv \nabla \times \mathbf{v} \\ &= \left( \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \right) \mathbf{i} + \left( \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} \right) \mathbf{j} + \left( \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right) \mathbf{k},\end{aligned}\quad (9.14)$$

where  $\mathbf{v} = (u, v, w)$ . Consequently,  $\boldsymbol{\omega}$  is twice the average angular velocity in the three coordinate planes. This also helps explain why  $\mathbf{W}$  in (8.67) is known as the vorticity tensor (see Exercise 9.8).

*Example (Plane Couette Flow)* As shown in Fig. 9.1, the fluid particles in Couette flow move in straight lines, and, consequently, appear to have no rotational component. However, using the solution (9.6) in (9.14), one obtains  $\boldsymbol{\omega} = (0, 0, -\dot{\gamma})$ . In other words, the vorticity is nonzero. To explain this, plane Couette flow can be thought of using traffic flow on a multilane road, where the fluid particles are the cars. This is shown in Fig. 9.7. The slowest lane is at  $y = 0$ , and the fastest lane is at  $y = h$ . Given a line of cars that start out at  $x = 0$ , after a short time they will have a linear distribution as shown in Fig. 9.7. A driver in one of middle lanes will see the car on the left a bit farther ahead, and the one on the right a bit farther behind. Hence, from the driver's perspective there has been a rotation in the orientation, the rotation being in the clockwise direction. This gives rise to a negative angular velocity, and this is why the  $z$ -component of the vorticity is negative for this flow. This example also shows that nonzero vorticity does not necessarily mean that the fluid particles themselves are rotating. The definition of vorticity assumes nothing about how the fluid particles interact, it only measures their respective orientations as they flow past each other. ■



**Fig. 9.7** Multilane traffic flow analogy used to explain vorticity in plane Couette flow

**Fig. 9.8** The rotational motion of a hurricane is an example of vortex type motion, with the eye containing the central axis



### 9.2.1 Vortex Motion

A vortex is a circular flow around a center, and is similar to what is seen in a tornado, hurricane, and in the swirling flow through a drain (Fig. 9.8). To study such motions, it is often convenient to use cylindrical coordinates, and the equations of motion in this coordinate system are given in Appendix E. The coordinates in this system are  $(r, \theta, z)$ , with corresponding velocity  $\mathbf{v} = (v_r, v_\theta, v_z)$ . Assuming the center of the vortex is the  $z$ -axis, then to have circular motion around the  $z$ -axis we assume that  $v_r = v_z = 0$ . This means there is no motion in either the  $z$ - or  $r$ -direction, and so the fluid particles move on circles centered on the  $z$ -axis. Making the additional assumption that  $v_\theta = v_\theta(r, t)$ , then the equations of motion reduce to

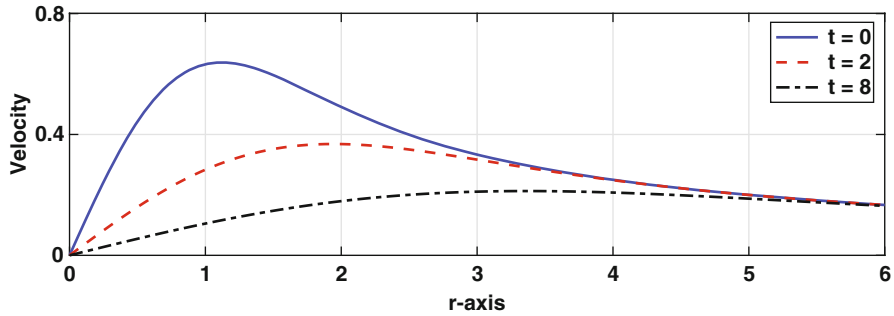
$$\frac{\partial v_\theta}{\partial t} = v \frac{\partial}{\partial r} \left( \frac{1}{r} \frac{\partial(r v_\theta)}{\partial r} \right), \quad (9.15)$$

$$\frac{\partial p}{\partial r} = \frac{\rho}{r} v_\theta^2. \quad (9.16)$$

This assumes the fluid is incompressible, and that there are no body forces. The vorticity for this flow, in cylindrical coordinates, is

$$\boldsymbol{\omega} = \left( 0, 0, \frac{1}{r} \frac{\partial(r v_\theta)}{\partial r} \right). \quad (9.17)$$

The momentum equation (9.15) is an old friend because it is the radially symmetric diffusion equation given in (4.82). In this case, the kinematic viscosity is the diffusion coefficient. The point source solution given in (4.85) gives rise to what is known as the Taylor vortex. The analysis of this vortex is carried out in Exercise 9.5, while we will investigate a related vortex in the following example.



**Fig. 9.9** Circumferential velocity (9.18) for a Oseen-Lamb vortex

*Example (Oseen-Lamb Vortex)* In this flow  $v_r = v_z = 0$ , and

$$v_\theta = \frac{\alpha}{r} \left( 1 - \exp\left(\frac{-r^2}{\beta^2 + 4vt}\right) \right). \quad (9.18)$$

It is not hard to show that this function satisfies (9.15) and is therefore an exact solution of the incompressible fluid equations. The pressure is found by integrating (9.16), and the vorticity is calculated using (9.17). One finds that

$$\omega = \left( 0, 0, \frac{2\alpha}{\beta^2 + 4vt} \exp\left(\frac{-r^2}{\beta^2 + 4vt}\right) \right). \quad (9.19)$$

The velocity (9.18) is shown in Fig. 9.9 at different time points, for  $\alpha = \beta = 4\nu = 1$ . This shows that the vortex is initially confined to the region near  $r = 0$ . As time passes the vortex slows down, with the maximum velocity moving outward from the center and decreasing in the process. This is due entirely to the viscosity of the fluid, and the result is that in the limit of  $t \rightarrow \infty$ , the vortex disappears. Also note that when this vortex starts out, there is a region near  $r = 0$  where there is little motion, which is reminiscent of the eye of the hurricane shown in Fig. 9.8. ■

### 9.3 Irrotational Flow

One of the ideas underlying the introduction of vorticity is that the motion of a fluid can be split into two components, a rotational part and a non-rotational part. How this can be done is obtained from the Helmholtz Representation Theorem, and this will be presented shortly. In preparation for this we introduce the concept of an irrotational flow.

**Irrotational Flow.** A fluid for which the vorticity is identically zero is said to be irrotational. The flow is rotational if the vorticity is nonzero anywhere in the flow.

So, not surprisingly given the formula for the vorticity in (9.19), an Oseen-Lamb vortex qualifies as rotational flow. Another non-surprising example is given next.

*Example (Rigid Body Rotation)* For rigid body motion around the  $z$ -axis, with angular velocity  $\omega$ , the rotation matrix  $\mathbf{Q}(t)$  is given in (8.14). In this case,  $\mathbf{v} = \mathbf{Q}'\mathbf{Q}^T \mathbf{x} = (-\omega y, \omega x, 0)^T$ , and so,  $\boldsymbol{\omega} = 2\omega \mathbf{k}$ . This is nonzero if  $\omega$  is nonzero, which means that merry-go-round motion is rotational. ■

One might guess that a flow which moves in a straight line is irrotational, but the plane Couette flow example shows that statement is incorrect. It is also incorrect to assume that if the flow is a vortex, then it must be rotational. The next example explains why.

*Example (Line Vortex)* In the special case of when  $v_r = v_z = 0$ , and  $v_\theta = v_\theta(r, t)$ , then the vorticity is given in (9.17). This will be zero if  $rv_\theta$  is constant. Consequently, an irrotational flow is achieved by taking

$$v_\theta = \frac{\alpha}{r}, \quad (9.20)$$

where  $\alpha$  is a constant. The flow in this case is circular motion around the  $z$ -axis, just as it is for the Oseen-Lamb vortex shown in Fig. 9.9. This is called a line vortex, and it produces irrotational flow. ■

As the above example clearly demonstrates, rotational motion around the origin does not necessarily mean that the vorticity is nonzero. The reason this is confusing is that vorticity is a local property of the flow, and it is determined by the relative movement of nearby fluid particles. This is not necessarily the same as what is happening to the flow on the macroscopic level. This is why the conclusion coming from the line vortex example, that this particular rotational flow around the origin is irrotational, is not self-contradictory.

One of the difficulties with assuming a flow is irrotational is that it is a statement about the absence of a property, namely no vorticity. The question arises whether it might be possible to characterize the solutions of the Navier-Stokes equation that are irrotational. To answer this, we will make use of the following result.

**Helmholtz Representation Theorem.** Assume  $\mathbf{q}(\mathbf{x})$  is a smooth function of  $\mathbf{x}$  in a domain  $D$ . In this case, there exists a scalar function  $\phi(\mathbf{x})$  and a vector function  $\mathbf{g}(\mathbf{x})$  so that, for  $\mathbf{x} \in D$ ,

$$\mathbf{q}(\mathbf{x}) = \nabla\phi + \nabla \times \mathbf{g}, \quad (9.21)$$

where  $\nabla \cdot \mathbf{g} = 0$ . The function  $\phi$  is called a scalar potential, and  $\mathbf{g}$  is a vector potential, for  $\mathbf{q}$ .

The proof of this theorem involves two vector identities and a result from partial differential equations. The first identity is that, given any smooth vector function  $\mathbf{h}(\mathbf{x})$ ,

$$\nabla^2 \mathbf{h} = \nabla(\nabla \cdot \mathbf{h}) - \nabla \times (\nabla \times \mathbf{h}).$$

The right-hand side of this equation resembles the result in (9.21), where  $\phi = \nabla \cdot \mathbf{h}$  and  $\mathbf{g} = -\nabla \times \mathbf{h}$ . What is needed is to find  $\mathbf{h}$  so that  $\nabla^2 \mathbf{h} = \mathbf{q}$ . This is known as Poisson's equation, and this is where the result from partial differential equations is needed. A particular solution of Poisson's equation is (Weinberger 1995)

$$\mathbf{h}(\mathbf{x}) = -\frac{1}{4\pi} \iiint_D \frac{\mathbf{q}(\mathbf{s})}{\|\mathbf{x} - \mathbf{s}\|} dV_s, \quad (9.22)$$

where the subscript  $s$  indicates integration with respect to  $\mathbf{s}$ . With this choice for  $\mathbf{h}$  we have derived an expression of the form given in (9.21). The only thing left to show is that  $\nabla \cdot \mathbf{g} = 0$ . This follows because  $\mathbf{g} = -\nabla \times \mathbf{h}$  and the vector identity that states, given any smooth vector function  $\mathbf{h}(\mathbf{x})$ ,  $\nabla \cdot (\nabla \times \mathbf{h}) = 0$ .

The above proof relied on the solution of Poisson's equation, and this requires certain conditions to be satisfied. If the closure of  $D$  is a bounded regular region, then the stated assumption that  $\mathbf{q}$  is smooth is sufficient. Specifically, what this means is that  $\nabla \times \mathbf{q}$  and  $\nabla \cdot \mathbf{q}$  are continuous functions. If  $D$  is not bounded, then the integral requires an additional condition, which is that  $\mathbf{q}$  goes to zero faster than  $\|\mathbf{x}\|^{-2}$  as  $\|\mathbf{x}\| \rightarrow \infty$ . It is, however, possible to modify the proof so this latter condition is not needed. The details concerning the extension to unbounded domains can be found in Gregory (1996).

It is worth pointing out that the proof is constructive in the sense that it provides formulas for the potential functions. It is not hard to show that, in the case of when  $D = \mathbb{R}^3$ , that (see Exercise 9.15)

$$\phi = \frac{1}{4\pi} \iiint \frac{\nabla_s \cdot \mathbf{q}}{\|\mathbf{x} - \mathbf{s}\|} dV_s, \quad (9.23)$$

and

$$\mathbf{g} = -\frac{1}{4\pi} \iiint \frac{\nabla_s \times \mathbf{q}}{\|\mathbf{x} - \mathbf{s}\|} dV_s. \quad (9.24)$$

Also, note that the representation concerns the spatial dependence of a function  $\mathbf{q}$ . If  $\mathbf{q}$  also depends on  $t$ , then the scalar and vector potential functions depend on  $t$ . Finally, the usefulness of the Helmholtz Representation Theorem is not limited to fluid dynamics, and a survey of its various applications can be found in Bhatia et al. (2013).



### 9.3.1 Potential Flow

According to the Helmholtz Representation Theorem, the velocity can be written as  $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ , where  $\mathbf{v}_1 = \nabla\phi$  and  $\mathbf{v}_2 = \nabla \times \mathbf{g}$ . If it turns out that  $\mathbf{v}_2 \equiv \mathbf{0}$ , then the flow is irrotational. So, the question is, if the flow is irrotational, then is it true that  $\mathbf{v}_2 \equiv \mathbf{0}$ ? The answer is yes, as shown in (9.24), in the case of when  $D = \mathbb{R}^3$ . For other domains the answer is yes or no, depending on what other assumptions are being made. Nevertheless, given the prominent role of such flows, we have the following definition.

**Potential Flow.** Any flow in which the velocity can be written as

$$\mathbf{v} = \nabla\phi \quad (9.25)$$

is called potential flow.

To investigate the mathematical consequences of this, we will assume that the fluid is incompressible and there are no body forces. The continuity equation  $\nabla \cdot \mathbf{v} = 0$  in this case reduces to

$$\nabla^2\phi = 0. \quad (9.26)$$

This means that the velocity can be found by simply solving Laplace's equation, and this is one of the reasons why potential flow is a centerpiece in most fluid dynamics textbooks. It is important to point out here that nothing has been said about the boundary conditions. These have major repercussions for potential flow, and this will be discussed in more detail shortly.

The pressure  $p$  for potential flow is determined by solving the momentum equation. In the  $x$ -direction, as given in (8.77), we have that

$$\rho \left( \frac{\partial^2\phi}{\partial x \partial t} + \frac{\partial\phi}{\partial x} \frac{\partial^2\phi}{\partial x^2} + \frac{\partial\phi}{\partial y} \frac{\partial^2\phi}{\partial x \partial y} + \frac{\partial\phi}{\partial z} \frac{\partial^2\phi}{\partial x \partial z} \right) = -\frac{\partial p}{\partial x} + \mu \nabla^2 \frac{\partial\phi}{\partial x}. \quad (9.27)$$

Given (9.26), then the viscous stress term  $\mu \nabla^2 \partial_x \phi$  in the above equation is zero. In other words, for irrotational flow the viscosity does not contribute to the momentum equation. With this, it is possible to rewrite (9.27) in the form

$$\frac{\partial}{\partial x} \left[ \frac{\partial\phi}{\partial t} + \frac{1}{2} \left( \frac{\partial\phi}{\partial x} \right)^2 + \frac{1}{2} \left( \frac{\partial\phi}{\partial y} \right)^2 + \frac{1}{2} \left( \frac{\partial\phi}{\partial z} \right)^2 + \frac{1}{\rho} p \right] = 0.$$

Not too surprisingly, the  $y$  and  $z$  momentum equations show that the  $y$  and  $z$  derivatives of the above quantity in the square brackets are zero. The conclusion is that

$$\frac{\partial\phi}{\partial t} + \frac{1}{2} \left( \frac{\partial\phi}{\partial x} \right)^2 + \frac{1}{2} \left( \frac{\partial\phi}{\partial y} \right)^2 + \frac{1}{2} \left( \frac{\partial\phi}{\partial z} \right)^2 + \frac{1}{\rho} p$$

is only a function of time. In other words,

$$p = p_0(t) - \rho \left[ \frac{\partial \phi}{\partial t} + \frac{1}{2} \left( \frac{\partial \phi}{\partial x} \right)^2 + \frac{1}{2} \left( \frac{\partial \phi}{\partial y} \right)^2 + \frac{1}{2} \left( \frac{\partial \phi}{\partial z} \right)^2 \right]. \quad (9.28)$$

Generalizing this to include a conservative body force, the following result is obtained (see Exercise 9.17).

**Bernoulli's Theorem.** *Assume the body force is conservative, so that  $\mathbf{f} = \nabla \Psi$ . In this case, if the flow is irrotational and incompressible, then*

$$p = p_0(t) - \rho \left( \frac{\partial \phi}{\partial t} + \frac{1}{2} \nabla \phi \cdot \nabla \phi - \Psi \right). \quad (9.29)$$

Consequently, once Laplace's equation is solved to find the potential function, (9.25) is used to find the velocity and (9.29) is used to find the pressure.

*Example (Line Vortex (Cont'd))* Using cylindrical coordinates, then  $\mathbf{v} = (v_r, v_\theta, v_z)$ , and

$$\nabla \phi = \left( \frac{\partial \phi}{\partial r}, \frac{1}{r} \frac{\partial \phi}{\partial \theta}, \frac{\partial \phi}{\partial z} \right). \quad (9.30)$$

To obtain  $v_r = v_z = 0$  it is required that  $\phi = \phi(\theta, t)$ . To have (9.20) hold it is required that  $\frac{\partial \phi}{\partial \theta} = \alpha$ . Therefore, the scalar potential function for the line vortex is  $\phi = \alpha \theta$ . The pressure, obtained from (9.29), is  $p = p_0 - \frac{1}{2} \rho \alpha^2 / r^2$ . ■

One question that has not been addressed is, how realistic is it to assume a flow is irrotational? In applications, in addition to the equations of motion, there are boundary and initial conditions, and these were conveniently ignored in the derivation of (9.23) and (9.24). The fact is that they can easily ruin the assumption of irrotationality. To explain why, consider the no-slip condition (8.6). This prescribes all three components of the velocity vector on the boundary. The equation to solve for an irrotational flow is Laplace's equation (9.26), from which the velocities are determined using (9.25). Mathematically, for Laplace's equation, one can only impose one condition on the boundary, and not three as required from the no-slip condition. The usual choice is to have the solution satisfy the impermeability condition (8.79). Therefore, if the flow is to be irrotational, the other two boundary conditions making up the no-slip condition would have to be selected to be consistent with the resulting solution of Laplace's equation. What this means is that irrotational flow in a viscous fluid is possible, but the boundary conditions have to be just right. An example is the line vortex above, where there are no boundaries, and hence no difficulties trying to satisfy the no-slip condition. Physically, what happens in most fluid problems is that the boundaries generate vorticity, which then spreads into the flow and causes it to be rotational. An example of this is shown in Fig. 9.10. One way to avoid this from happening, in addition to adjusting the

**Fig. 9.10** The motion of the airflow around a plane generates vorticity into the flow (Morris 2006). This is evident in the motion of the clouds behind the plane in the photograph



boundary conditions, is to assume the fluid viscosity is zero. This produces what is known as an inviscid fluid, and this is the subject of the next section.

Because of the complications of the no-slip condition, most textbooks associate potential flow with an inviscid fluid. In fact, to overcome this association, in the research literature the above discussion would be referred to as viscous potential flow, just to make sure to point out that the viscosity has not been assumed to be zero. Those interested in learning about some of the consequences of keeping the viscosity in a potential flow should consult Joseph 2006.

## 9.4 Ideal Fluid

As seen in Table 9.1, the viscosity of air is much less than it is, for example, for water. It is for this reason that when studying air flow that it is often assumed to be inviscid, which means the viscosity is zero. If, in addition, the fluid is assumed to be incompressible, then one has what is called an *ideal fluid*. The equations of motion in this case are

$$\rho(\partial_t + \mathbf{v} \cdot \nabla)\mathbf{v} = -\nabla p + \rho \mathbf{f}, \quad (9.31)$$

$$\nabla \cdot \mathbf{v} = 0. \quad (9.32)$$

The above system is known as the Euler equations. The absence of viscosity means that the no-slip condition is inappropriate, but the impermeability boundary condition still applies.

It is assumed in what follows that there is a unique solution of the ideal fluid problem and it is smooth. As it turns out, this is not always true. The reason is that the problem is similar, in terms of its mathematical properties, to the nonlinear traffic problem considered in Chap. 5. As we discovered for traffic flow, it is very easy to generate non-smooth solutions, and to obtain a unique solution we had to introduce

something called an admissibility condition. In the examples to be considered below these complications do not arise, but this issue will be discussed again at the end of this section.

*Example (Plane Couette Flow Revisited)* The lack of viscosity has some interesting consequences. As an example, suppose in the plane Couette flow problem that the fluid is not moving at  $t = 0$ . With a viscous fluid, because of the no-slip condition, when the upper surface starts to move it pulls the nearby fluid with it. After a short amount of time the fluid between the two plates approaches a steady flow, and the solution given in (9.6) applies. This does not happen if the fluid is inviscid. The only boundary condition at  $y = h$  is the impermeability condition, which is that the velocity in the vertical direction is zero. The motion of the plate has no effect on the fluid, and so the fluid remains at rest. Therefore, the solution of the plane Couette flow problem for an ideal fluid is simply  $\mathbf{v} = \mathbf{0}$  and  $p = p_0$ . ■

### 9.4.1 Circulation and Vorticity

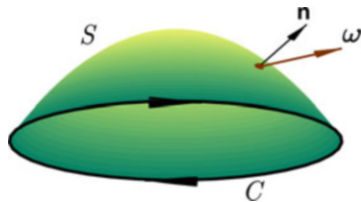
An important property of an ideal fluid is that if it starts out irrotational, and the body force is conservative, the flow is irrotational for all time. To explain why, we start with the surface integral

$$\iint_S \boldsymbol{\omega} \cdot \mathbf{n} dA, \quad (9.33)$$

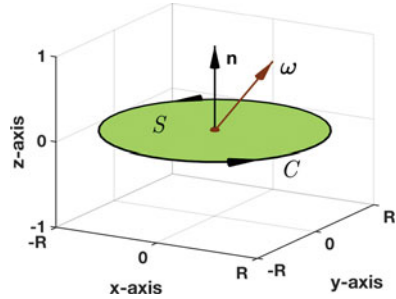
where  $S$  is an oriented smooth surface that is bounded by a simple, closed, smooth boundary curve  $C$  with positive orientation (see Fig. 9.11). The integral gives the total of the vorticity, in the normal direction, over the surface  $S$ . The first step is to recall Stokes' theorem, which states that

$$\iint_S (\nabla \times \mathbf{v}) \cdot \mathbf{n} dA = \int_C \mathbf{v} \cdot d\mathbf{x}.$$

**Fig. 9.11** Example of the surface  $S$ , and bounding curve  $C$ , used in the definition of circulation



**Fig. 9.12** Surface  $S$ , and contour  $C$ , used to calculate the circulation for an Oseen-Lamb vortex



Since  $\omega = \nabla \times \mathbf{v}$ , then from (9.33) we have

$$\iint_S \omega \cdot \mathbf{n} dA = \int_C \mathbf{v} \cdot d\mathbf{x}. \quad (9.34)$$

It is the last integral that we will work with, and so let

$$\Gamma(t) = \int_C \mathbf{v} \cdot d\mathbf{x}. \quad (9.35)$$

The function  $\Gamma(t)$  is called the *circulation*.

*Example (Oseen-Lamb Vortex Revisited)* The vorticity for the Oseen-Lamb vortex is given in (9.19). Suppose we want to calculate the circulation when the curve  $C$  is the circle in the  $x, y$ -plane, with radius  $R$  and centered at the origin (see Fig. 9.12). It is easier, in this case, to use (9.34) and write

$$\Gamma(t) = \iint_S \omega \cdot \mathbf{n} dA.$$

Using cylindrical coordinates, the surface  $S$  is the disk  $r \leq R$  in the plane  $z = 0$ , and  $\mathbf{n} = (0, 0, 1)$ . From (9.19),  $\omega \cdot \mathbf{n} = 2\alpha q(t) \exp(-q(t)r^2)$ , where  $q(t) = 1/(\beta^2 + 4vt)$ , and so

$$\begin{aligned} \Gamma &= \int_0^{2\pi} \int_0^R 2\alpha q(t) \exp(-q(t)r^2) r dr d\theta \\ &= 2\pi\alpha [1 - \exp(-q(t)R^2)]. \end{aligned}$$

Consequently, the circulation starts out with the value

$$\Gamma_0 = 2\pi\alpha [1 - \exp(-R^2/\beta^2)],$$

and then decays to zero as  $t \rightarrow \infty$ . In contrast, for an ideal fluid, so  $\nu = 0$ , the circulation has the constant value  $\Gamma_0$ . It is this property, that the circulation is constant for an ideal fluid, that is the central idea of the theorem given below. ■

Before stating the theorem, the concept of a material curve needs to be explained. Suppose one starts out, at  $t = 0$ , with a simple closed curve. As time progresses, the material points making up this initial curve move with the fluid, deforming the original shape. Due to the impenetrability of matter assumption, the points never intersect, so the shape remains a simple closed curve. This is what is known as a material curve. With this, we can now state a well-known result about the circulation.

**Kelvin's Circulation Theorem.** *Assuming that the body force is conservative, which means that  $\mathbf{f} = \nabla\Psi$ , then for an ideal fluid, if  $C$  is a material curve, then  $\frac{d\Gamma}{dt} = 0$ .*

To prove this, for the no body force case, we need what effectively is a Reynolds transport theorem for line integrals. The first step is to use material coordinates to get the time dependence out of the limits of integration, and so, at  $t = 0$  assume the curve  $C$  is given as  $\mathbf{A} = \mathbf{G}(s)$ , for  $a \leq s \leq b$ . At later times the curve is described as  $\mathbf{x} = \mathbf{X}(\mathbf{G}(s), t)$ . With this,  $d\mathbf{x} = \mathbf{F}\mathbf{G}'(s)ds$ , where  $\mathbf{F}$  is the deformation gradient given in (8.10). So (9.35) becomes

$$\Gamma = \int_a^b \mathbf{V} \cdot \mathbf{F}\mathbf{G}'(s)ds,$$

where  $\mathbf{V}$  and  $\mathbf{F}$  are evaluated at  $\mathbf{A} = \mathbf{G}$ . Taking the time derivative yields

$$\frac{d\Gamma}{dt} = \int_a^b \left( \frac{\partial \mathbf{V}}{\partial t} \cdot \mathbf{F}\mathbf{G}'(s) + \mathbf{V} \cdot \frac{\partial \mathbf{F}}{\partial t} \mathbf{G}'(s) \right) ds. \quad (9.36)$$

Using the results from Exercise 8.7, and remembering that  $\mathbf{V}$  is evaluated at  $\mathbf{A} = \mathbf{G}$ , it follows that

$$\begin{aligned} \mathbf{V} \cdot \frac{\partial \mathbf{F}}{\partial t} \mathbf{G}'(s) &= \mathbf{V} \cdot \nabla_{\mathbf{A}} \mathbf{V}\mathbf{G}'(s) \\ &= \frac{d}{ds} \frac{1}{2} (\mathbf{V} \cdot \mathbf{V}). \end{aligned}$$

Given that the curve is closed, then (9.36) reduces to

$$\begin{aligned} \frac{d\Gamma}{dt} &= \int_a^b \frac{\partial \mathbf{V}}{\partial t} \cdot \mathbf{F}\mathbf{G}'(s)ds \\ &= \int_C \frac{D\mathbf{v}}{Dt} \cdot d\mathbf{x}. \end{aligned} \quad (9.37)$$

What remains is to recall a property of line integrals. Specifically, given any smooth function  $\phi$ , and a closed curve  $C$ , the following holds  $\int_C \nabla \phi \cdot d\mathbf{x} = 0$ . From (9.31) we have that  $\frac{D\mathbf{v}}{Dt} = -\frac{1}{\rho} \nabla p$ , where  $\rho$  is constant. Therefore, from (9.37), we have that  $\frac{d\Gamma}{dt} = 0$ . The case of when the body force is not zero is outlined in Exercise 9.17.

The above result will enable us to make the stated conclusion about the irrotationality of an ideal fluid.

**Helmholtz's Third Vorticity Theorem.** *If an ideal fluid, with a conservative body force, is irrotational at  $t = 0$ , then it is irrotational for all time.*

The proof of this starts with using Stokes' theorem to write the circulation as

$$\Gamma = \iint_S \boldsymbol{\omega} \cdot \mathbf{n} dA.$$

This shows that because  $\boldsymbol{\omega} = \mathbf{0}$  at  $t = 0$ , then  $\Gamma = 0$  at  $t = 0$ . Given that  $\Gamma$  is constant, it follows that  $\Gamma = 0$  for all time. To use this observation to prove the vorticity is always zero, suppose  $\boldsymbol{\omega}$  is nonzero at some point in the flow. In this case, given that  $\boldsymbol{\omega}$  is continuous, it is possible to find a small surface containing this point for which the above integral is nonzero. This is a contradiction, and therefore  $\boldsymbol{\omega}$  must be zero everywhere.

As an example, the above theorem shows that an ideal fluid which starts at rest is irrotational for all time. The reason is that because  $\mathbf{v} = \mathbf{0}$  at  $t = 0$ , then  $\boldsymbol{\omega} = \mathbf{0}$  at  $t = 0$ .

### 9.4.2 Potential Flow

What we have been able to show is that if an ideal fluid is irrotational at  $t = 0$ , then it is possible to introduce a potential function  $\phi$  so that

$$\mathbf{v} = \nabla \phi, \quad (9.38)$$

and

$$p = p_0(t) - \rho \left( \frac{\partial \phi}{\partial t} + \frac{1}{2} \nabla \phi \cdot \nabla \phi \right). \quad (9.39)$$

To find  $\phi$  one solves

$$\nabla^2 \phi = 0, \quad (9.40)$$

along with the appropriate boundary conditions. For example, if the equation is to be solved in a bounded domain  $D$ , then an often used boundary condition is to prescribe the normal velocity  $v_n$  on the boundary  $\partial D$ . This gives the condition

$$\frac{\partial \phi}{\partial n} = v_n \quad \text{on } \partial D, \quad (9.41)$$

or equivalently

$$\nabla \phi \cdot \mathbf{n} = v_n \quad \text{on } \partial D. \quad (9.42)$$

For consistency with the assumption of incompressibility, it is required that

$$\iint_{\partial D} v_n \, dA = 0.$$

If the problem involves a pressure boundary condition, then the corresponding boundary condition for  $\phi$  is obtained using (9.39). However, this can make the problem harder to solve because the  $\nabla \phi \cdot \nabla \phi$  term results in the problem being nonlinear.

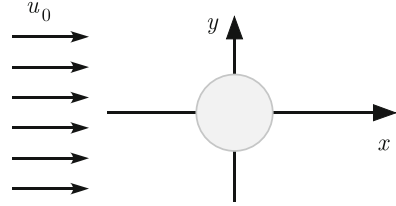
Any fluid flow in which the velocity satisfies (9.38) is known as potential flow. Although this might seem obvious, it differs from the definition used in some textbooks on fluid dynamics, where a potential flow is defined as “an irrotational flow in an inviscid and incompressible fluid.” The reason for including these additional qualifications, as explained at the end of Sect. 9.3.1, is the difficulty of obtaining a potential flow when the fluid is viscous. However, it is inappropriate to include them. The reason is that potential flow is a statement about a fluid’s motion, while the statement that it is inviscid is an assumption about its material properties.

We have been making a series of simplifying assumptions in this chapter, attempting to obtain a more tractable mathematical problem. By this measure, we have been extraordinarily successful because we have reduced a coupled system of nonlinear equations down to the single linear equation in (9.40). This has been done by excluding the effects of viscosity, and assuming the flow is irrotational. This degree of simplification helps explain the interest in potential flow. It is also why textbooks on the applications of complex variables inevitably have a chapter on fluid flow, although they must limit their analysis to flow in two dimensions. The question is, however, just how realistic is it to assume a potential flow? The next example will shed some light on this topic.

*Example (Potential Flow Past a Cylinder)* Consider air flow over a solid cylinder of radius  $R$  centered on the  $z$ -axis, as shown in Fig. 9.13. It is assumed that the flow is from left to right, and the specific condition is that  $\mathbf{v} = (u_0, 0, 0)$  as  $x \rightarrow -\infty$ . The flow must also satisfy the impermeability condition on the surface of the cylinder, and this means that  $\mathbf{v} \cdot \mathbf{n} = 0$  on  $|\mathbf{x}| = R$ , where  $\mathbf{n}$  is the unit normal to the boundary. Given the geometry and flow at infinity it is reasonable to expect there is



**Fig. 9.13** Cross-section for uniform flow past a cylinder



no flow in the  $z$ -direction, and the potential function is independent of  $z$ . With this, Laplace's equation in cylindrical coordinates becomes

$$\frac{\partial^2 \phi}{\partial r^2} + \frac{1}{r} \frac{\partial \phi}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \phi}{\partial \theta^2} = 0, \text{ for } r > R. \quad (9.43)$$

The impermeability condition (9.41) takes the form

$$\frac{\partial \phi}{\partial r} = 0, \text{ for } r = R, \quad (9.44)$$

and the flow at infinity requires that

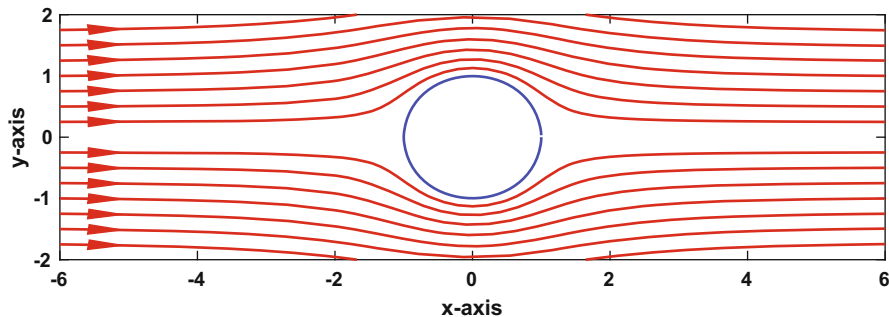
$$\left. \begin{aligned} \cos \theta \frac{\partial \phi}{\partial r} + \sin \theta \frac{1}{r} \frac{\partial \phi}{\partial \theta} &= u_0, \\ \sin \theta \frac{\partial \phi}{\partial r} + \cos \theta \frac{1}{r} \frac{\partial \phi}{\partial \theta} &= 0. \end{aligned} \right\} \text{ for } r \rightarrow \infty \quad (9.45)$$

This is one of the few unbounded domain problems for which the method of separation of variables can be used. So, assuming that  $\phi(r, \theta) = F(r)G(\theta)$  one finds from (9.43) that  $F = \alpha r^n + \beta r^{-n}$  and  $G = A \cos(n\theta) + B \sin(n\theta)$ . Because the solution must be  $2\pi$  periodic in  $\theta$ , it is required that  $n$  be a positive integer. From (9.44) it follows that  $\beta = \alpha R^{2n}$ . Imposing (9.45) yields  $n = 1$ ,  $\alpha A = u_0$ , and  $B = 0$ . The resulting potential is

$$\phi = u_0 \cos \theta \left( r + \frac{R^2}{r} \right). \quad (9.46)$$

The velocity field is therefore

$$\begin{aligned} v_r &= \frac{\partial \phi}{\partial r} = u_0 \cos \theta \left( 1 - \frac{R^2}{r^2} \right), \\ v_\theta &= \frac{1}{r} \frac{\partial \phi}{\partial \theta} = -u_0 \sin \theta \left( 1 + \frac{R^2}{r^2} \right). \end{aligned}$$



**Fig. 9.14** Flow around a cylinder, calculated using the potential function given in (9.46). In this calculation,  $u_0 = R = 1$

How to use the formula for the velocity to determine the paths of individual fluid particles, what are called pathlines, is explained in Exercise 9.12(d). This is easily done numerically, and the results from this calculation are shown in Fig. 9.14. ■

If air can be assumed to be an ideal fluid, then it would seem that potential flow could be used in aerodynamics to help understand flight. As an example, you could think of Fig. 9.14 as the flow around an airplane wing that has a circular cross-section. You also might think that this is not particularly realistic because cross-sections of airplane wings are relatively thin, to help reduce the drag and increase lift. Well, let's see about this. The pressure is determined by substituting (9.46) into (9.39), yielding

$$p = \frac{1}{2}\rho \left( \frac{u_0 R}{r} \right)^2 \left( 4 \cos^2 \theta - 2 - \frac{R^2}{r^2} \right). \quad (9.47)$$

The force on the circular cross-section is

$$\mathbf{F} = - \int_C p \mathbf{n} ds,$$

where  $C$  is the boundary circle  $x^2 + y^2 = R^2$ , and  $\mathbf{n}$  is the unit outward normal to the circle. The  $x$  and  $y$  components of this force are

$$F_x = -R \int_0^{2\pi} p(R, \theta) \cos(\theta) d\theta,$$

$$F_y = -R \int_0^{2\pi} p(R, \theta) \sin(\theta) d\theta.$$

A straightforward calculation shows that both integrals are zero. In other words, the drag  $F_x$ , and the lift  $F_y$ , are both zero. As it turns out, this happens with any shape,

as long as the fluid is ideal and the flow is steady and irrotational. This is clearly at odds with what is expected, and it is known as *d'Alembert's paradox*. It is possible to produce lift, an essential requirement to be able to fly, if the flow is rotational. This result is known as the Kutta-Joukowski theorem, but as we saw earlier, it is impossible to get an ideal fluid to be rotational if you start from rest. In other words, if you strap on a pair of wings and starting running in still air, there is no way you are taking off, no matter how fast you are able to run. What this means is that if you want your airplane to fly it is essential that the fluid is viscous. Or, more precisely, that the contribution of the viscosity in generating vorticity from the solid surface of the wing is accounted for in the model. One method how this can be done is explained in the next section.

### 9.4.3 End Notes

An important issue that arises when assuming the fluid is inviscid concerns the regularity of the solution. Viscosity usually acts to smooth out jumps and other irregular behavior. By not having viscosity, we have equations similar to those used to model traffic flow. This means that shock wave solutions are possible, and the uniqueness of the solution is an issue. In traffic flow we introduced an admissibility condition to determine uniqueness, but for multidimensional fluid problems there are still questions related to the appropriate condition. In fact, there are several open problems associated with the Euler equations. One that has generated considerable interest is the Euler blow-up problem. It is suspected that the solution of the three-dimensional Euler equations develops a singularity in finite time, but no one has been able to prove this assertion. This means that most of the evidence has come from numerical solutions, but even this has been contradictory. An interesting survey of the blow-up problem, as well as other aspects of the Euler equations, can be found in the proceedings of the conference, Euler Equations: 250 Years On (Eyink et al. 2008).

## 9.5 Boundary Layers

The assumption that a fluid is inviscid corresponds to setting the viscosity equal to zero in the Navier-Stokes equation (8.77). What drops out of the equation in this case is the highest spatial derivative in the problem. As we found in Sect. 2.5, this is the type of limit that is associated with the appearance of a boundary layer. The fact that boundary layers might occur in the flow of a viscous fluid is not surprising given the rapid transitions shown in Fig. 8.6. However, the situation is not as straightforward as what occurred in Chap. 2, because the viscous fluid problem is time dependent, and it is not clear what exactly the assumption “small viscosity”

means. To get started, we will consider an example that illustrates what happens in a time-dependent flow.

### 9.5.1 Impulsive Plate

This example is known as Stokes' first problem, and it is one of the few time dependent solutions known for the Navier-Stokes equation. It is assumed that the fluid is incompressible, has no body forces, and it occupies the region  $y > 0$ . Also, it is at rest for  $t < 0$ , and at  $t = 0$  the lower boundary, at  $y = 0$ , is given the constant velocity  $\mathbf{v} = (u_0, 0, 0)$ . This situation is similar to the plane Couette flow problem, in the sense that a planar boundary surface produces a flow in the  $x$ -direction. For this reason, the argument used to solve the Couette flow problem can be used here. Assuming that  $\mathbf{v} = (u(y, t), 0, 0)$ , then the problem reduces to solving

$$\begin{aligned}\rho(u_t + u\partial_x u) &= -\partial_x p + \mu\partial_y^2 u, \\ 0 &= -\partial_y p, \\ \partial_x u &= 0.\end{aligned}$$

As before, it follows that  $p$  is constant, and the entire problem reduces to solving

$$\frac{\partial u}{\partial t} = \nu \frac{\partial^2 u}{\partial y^2}, \quad (9.48)$$

where  $u(y, 0) = 0$ ,  $u(0, t) = u_0$ , and  $u(\infty, t) = 0$ . Also,  $\nu = \mu/\rho$  is the kinematic viscosity. This diffusion problem was solved in Sect. 1.4 using a similarity variable. The solution, given in (1.63), is

$$u(y, t) = u_0 \operatorname{erfc}\left(\frac{y}{2\sqrt{\nu t}}\right), \quad (9.49)$$

where

$$\operatorname{erfc}(\eta) = 1 - \frac{2}{\sqrt{\pi}} \int_0^\eta e^{-s^2} ds. \quad (9.50)$$

This solution is shown in Fig. 9.15 at three time values, in the case of when  $u_0 = \nu = 1$ . What is seen is that the effect of the moving plate is limited to the region near  $y = 0$ , which is expected of a boundary layer. However, as time passes the effects spread through the fluid domain, and this is due to the diffusive nature of the viscous stress. This gives rise to what is known as a diffusive boundary layer. To quantify what this means, in the engineering literature the boundary layer thickness is defined to be the distance between the boundary and the point where the velocity is 1% of

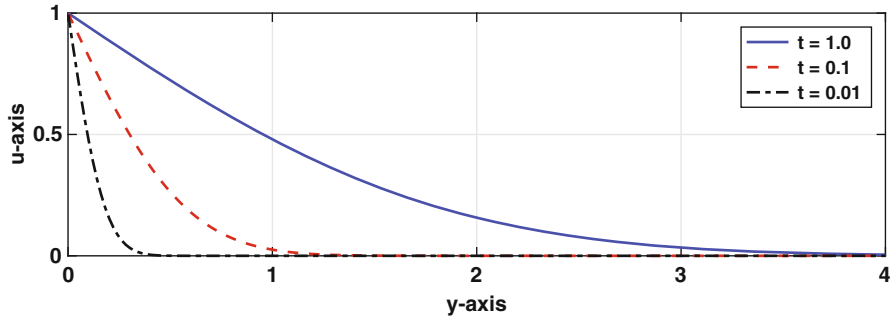


Fig. 9.15 Solution (9.49) of the impulsive plate problem, at three time values

the imposed value. Given that  $\operatorname{erfc}(\eta) = 0.01$  for  $\eta \approx 1.8$ , then the boundary layer thickness in this problem is approximately  $y = 3.6\sqrt{\nu t}$ . Consequently, this layer grows and spreads through the fluid region.

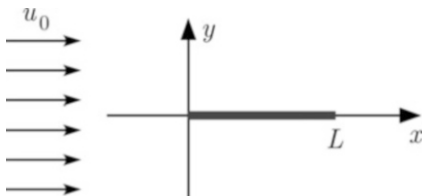
The existence of a boundary layer separates the flow into an inner and outer region. In the outer region the fluid can be approximated to be inviscid, and the viscous effects are confined to the inner, or boundary layer, region. This observation is routinely used in the numerical solution of the Navier-Stokes equation, because the resolution needed in the inviscid region is usually much less than what is needed in the boundary layer. This is seen in Fig. 2.16, where the grid structure near the surface of the plane is much finer than the one used in the outer, inviscid, flow region.

### 9.5.2 Blasius Boundary Layer

The boundary layer example to be considered involves the steady flow over a stationary flat plate (see Fig. 9.16). The plate occupies the plane  $y = 0$ , for  $0 < x < L$ , and the flow is coming in from the left. Assuming that the flow is steady, and that there is no flow in the  $z$ -direction, then the fluid equations are

$$\begin{aligned}\rho \left( u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} \right) &= -\frac{\partial p}{\partial x} + \mu \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right), \\ \rho \left( u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} \right) &= -\frac{\partial p}{\partial y} + \mu \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right), \\ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} &= 0.\end{aligned}$$

**Fig. 9.16** Flow over a flat plate used to study viscous boundary layers



The boundary conditions are

$$\mathbf{v} = (0, 0, 0) \quad \text{on } y = 0, 0 < x < L,$$

$$\mathbf{v} = (u_0, 0, 0) \quad \text{for } y \rightarrow -\infty.$$

To undertake a boundary layer analysis we must nondimensionalize the problem. This is done by letting  $x = L\bar{x}$ ,  $y = L\bar{y}$ ,  $u = u_0\bar{u}$ ,  $v = u_0\bar{v}$ , and  $p = p_c\bar{p}$ , where  $p_c = \rho u_0^2$ . In this case the equations of motion become

$$\bar{u} \frac{\partial \bar{u}}{\partial \bar{x}} + \bar{v} \frac{\partial \bar{u}}{\partial \bar{y}} = -\frac{\partial \bar{p}}{\partial \bar{x}} + \varepsilon^2 \left( \frac{\partial^2 \bar{u}}{\partial \bar{x}^2} + \frac{\partial^2 \bar{u}}{\partial \bar{y}^2} \right), \quad (9.51)$$

$$\bar{u} \frac{\partial \bar{v}}{\partial \bar{x}} + \bar{v} \frac{\partial \bar{v}}{\partial \bar{y}} = -\frac{\partial \bar{p}}{\partial \bar{y}} + \varepsilon^2 \left( \frac{\partial^2 \bar{v}}{\partial \bar{x}^2} + \frac{\partial^2 \bar{v}}{\partial \bar{y}^2} \right), \quad (9.52)$$

$$\frac{\partial \bar{u}}{\partial \bar{x}} + \frac{\partial \bar{v}}{\partial \bar{y}} = 0, \quad (9.53)$$

where

$$\varepsilon^2 = \frac{\mu}{\rho L u_0}. \quad (9.54)$$

From (1.20), we have that  $\varepsilon^2 = 1/Re$ , where  $Re$  is Reynolds number for the flow. So,  $\varepsilon^2$  is the inverse of the Reynolds number. Our assumption that the viscosity is small translates into the assumption that the Reynolds number is large. As an example, consider the flow over an airplane wing. The width of the wing on the Boeing 787 is 18 ft (5.5 m) and cruises at a speed of 561 mph (903 km/h). In this case,  $Re = 4 \times 10^7$ , which certainly qualifies as high Reynolds number flow.

The reduction of the above problem will closely follow the format used in Sect. 2.5, although the calculations are a bit more involved.

### Outer Solution

The expansion in this region is assumed to have the form  $\bar{\mathbf{v}} \sim \bar{\mathbf{v}}_0 + \varepsilon \bar{\mathbf{v}}_1 + \dots$  and  $\bar{p} \sim \bar{p}_0 + \varepsilon \bar{p}_1 + \dots$ . The problem for the first term, obtained by setting  $\varepsilon = 0$  in (9.51) - (9.53), is the problem for an inviscid flow. The solution is just  $\mathbf{v}_0 = (u_0, 0, 0)$ , and  $\bar{p}_0$  is a constant. It is assumed, for simplicity, that  $\bar{p}_0 = 0$ .

*Boundary Layer Solution*

The boundary layer coordinate is

$$Y = \frac{\bar{y}}{\varepsilon}.$$

As in Sect. (2.5), capitals will be used to designate the dependent variables in the boundary layer region. With this, (9.51)–(9.53) take the form

$$\bar{U} \frac{\partial \bar{U}}{\partial \bar{x}} + \frac{1}{\varepsilon} \bar{V} \frac{\partial \bar{U}}{\partial Y} = -\frac{\partial \bar{P}}{\partial \bar{x}} + \varepsilon^2 \frac{\partial^2 \bar{U}}{\partial \bar{x}^2} + \frac{\partial^2 \bar{U}}{\partial Y^2}, \quad (9.55)$$

$$\bar{U} \frac{\partial \bar{V}}{\partial \bar{x}} + \frac{1}{\varepsilon} \bar{V} \frac{\partial \bar{V}}{\partial Y} = -\frac{1}{\varepsilon} \frac{\partial \bar{P}}{\partial Y} + \varepsilon^2 \frac{\partial^2 \bar{V}}{\partial \bar{x}^2} + \frac{\partial^2 \bar{V}}{\partial Y^2}, \quad (9.56)$$

$$\frac{\partial \bar{U}}{\partial \bar{x}} + \frac{1}{\varepsilon} \frac{\partial \bar{V}}{\partial Y} = 0, \quad (9.57)$$

The appropriate expansions in this case are  $\bar{U} \sim \bar{U}_0 + \dots$ ,  $\bar{V} \sim \varepsilon(\bar{V}_0 + \dots)$ , and  $\bar{P} \sim \bar{P}_0 + \dots$ . Introducing these into (9.55)–(9.57), and letting  $\varepsilon \rightarrow 0$  we obtain

$$\bar{U}_0 \frac{\partial \bar{U}_0}{\partial \bar{x}} + \bar{V}_0 \frac{\partial \bar{U}_0}{\partial Y} = -\frac{\partial \bar{P}_0}{\partial \bar{x}} + \frac{\partial^2 \bar{U}_0}{\partial Y^2}, \quad (9.58)$$

$$\frac{\partial \bar{P}_0}{\partial Y} = 0, \quad (9.59)$$

$$\frac{\partial \bar{U}_0}{\partial \bar{x}} + \frac{\partial \bar{V}_0}{\partial Y} = 0. \quad (9.60)$$

From the no-slip condition on the plate, it is required that

$$(\bar{U}_0, \bar{V}_0) = (0, 0) \quad \text{on } Y = 0, \quad 0 < \bar{x} < 1. \quad (9.61)$$

Moreover, the solution must match with the outer solution, and for this reason it is required that

$$\bar{U}_0 \rightarrow 1 \quad \text{and} \quad \bar{P}_0 \rightarrow 0 \quad \text{as } Y \rightarrow \infty, \quad \text{for } 0 < \bar{x} < 1. \quad (9.62)$$

There is a matching condition for  $\bar{V}_0$ , but it is not needed at the moment and this will be explained after the solution is derived.

From (9.59) and (9.62) it follows that  $\bar{P}_0 = 0$ . The usual method for finding the velocity functions is to introduce a stream function  $\psi(\bar{x}, Y)$ , which is defined so that

$$\bar{U}_0 = \frac{\partial \psi}{\partial Y}, \quad (9.63)$$

$$\bar{V}_0 = -\frac{\partial \psi}{\partial \bar{x}}. \quad (9.64)$$

By doing this, the continuity equation (9.57) is satisfied automatically. This leaves the momentum equation (9.55), which reduces to

$$\frac{\partial \psi}{\partial Y} \frac{\partial^2 \psi}{\partial Y \partial \bar{x}} - \frac{\partial \psi}{\partial \bar{x}} \frac{\partial^2 \psi}{\partial Y^2} = \frac{\partial^3 \psi}{\partial Y^3}. \quad (9.65)$$

The boundary (9.61) and matching (9.62) conditions transform into the following:

$$\frac{\partial \psi}{\partial Y} = \frac{\partial \psi}{\partial \bar{x}} = 0, \quad \text{on } Y = 0, \quad (9.66)$$

and

$$\frac{\partial \psi}{\partial Y} \rightarrow 1, \quad \text{as } Y \rightarrow \infty. \quad (9.67)$$

As an aside, something that was not explained above is where the idea of using a stream function comes from. The answer is the Helmholtz Representation Theorem (9.21). When the flow is incompressible and two dimensional as in the present example, then the velocity vector can be written as  $\mathbf{v} = \nabla \times \mathbf{g}$ , where  $\mathbf{g} = (0, 0, \psi)$ . Expanding the curl, one obtains  $\mathbf{v} = (\partial_y \psi, -\partial_x \psi, 0)$ , and this gives rise to the stream function.

It is not possible to find an analytical solution of the above problem for the stream function. However, it is possible to come close if we make one more assumption. Instead of a plate of finite length, we assume that the plate is semi-infinite and occupies the interval  $0 \leq \bar{x} < \infty$ . This gives rise to what is known as the Blasius boundary layer problem, and it can be reduced by introducing a similarity variable. Specifically, assuming that  $\psi = \sqrt{\bar{x}} f(\eta)$ , where  $\eta = Y/\sqrt{\bar{x}}$ , then (9.65) reduces to

$$f''' + \frac{1}{2} f f'' = 0, \quad \text{for } 0 < \eta < \infty, \quad (9.68)$$

where (9.66) and (9.67) become

$$f(0) = f'(0) = 0, \quad \text{and} \quad f'(\infty) = 1. \quad (9.69)$$



One might argue that we have not made much progress, because the solution of the above problem is not known. However, the ordinary differential equation (9.68) is simpler than the partial differential equation (9.65), and this does provide some benefit. For example, it is much easier to solve (9.68) numerically than it is to solve (9.65) numerically.

Before working out an example, one last comment to make is that once the function  $f(\eta)$  is determined, then the velocity functions are calculated using the formulas

$$\bar{U}_0 = f'(\eta), \quad (9.70)$$

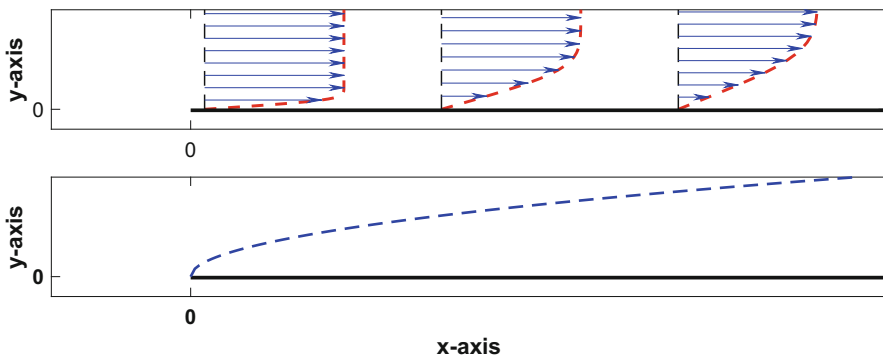
$$\bar{V}_0 = -\frac{1}{2\sqrt{x}}(f - \eta f'). \quad (9.71)$$

These expressions are obtained by substituting the similarity solution into (9.63) and (9.64).

*Example (Numerical Solution)* To use a numerical method to solve (9.68) it is a bit easier to rewrite the equation as a system by letting  $g = f'$ . In this case the equation can be written as

$$\begin{aligned} f' &= g, \\ g'' &= -\frac{1}{2}fg'. \end{aligned}$$

The boundary conditions (9.69) become  $g(0) = 0$ ,  $g(\infty) = 1$ , and  $f(0) = 0$ . With this, it is relatively straightforward to use finite differences to solve the problem (Holmes 2005). The result of such a calculation is given in Fig. 9.17. The upper

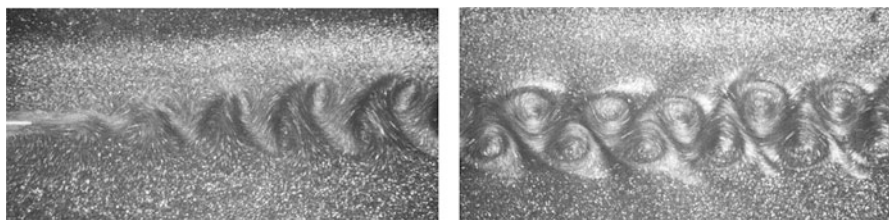


**Fig. 9.17** Flow over a flat plate, as determined from solving (9.68) and (9.69). The upper graph shows  $\bar{U}_0$ , as a function of  $Y$ , at three points on the plate. The dashed curve in the lower graph is where  $\bar{U}_0 = 0.99$

graph shows the horizontal velocity  $u$  at three locations along the plate. As required, the velocity is zero on the plate, and as the vertical distance from the plate increases it approaches the constant velocity of the outer region. It is also evident that the velocity reaches this constant value fairly quickly for a point on the plate that is near the leading edge, where  $\bar{x} = 0$ , and less so as the distance from the leading edge increases. The reason is that the boundary layer on the plate grows with distance from the leading edge. Using the engineering definition that the boundary layer thickness is where the flow reaches 99% of the outer flow value, the dashed curve shown in the lower graph is obtained. The shape of this curve can be explained using (9.71). By definition, the dashed curve is where  $\bar{U}_0 = 0.99$ , and this means that  $f'(\eta) = 0.99$ . Letting the solution of this equation be  $\eta_0$ , then, because  $\eta = Y/\sqrt{\bar{x}}$ , we have that the dashed curve is  $Y = \eta_0\sqrt{\bar{x}}$ . ■

The above example illustrates how a flow can be separated into an outer, inviscid, region, and a boundary layer where the viscous effects play an important role. This requires a large value for the Reynolds number. It is also based on the solution for an infinitely long plate, something that is rather rare in the real world. When the plate has finite length, a wake is formed downstream from the plate. An example of this is shown in Fig. 9.18. The pattern seen in the wake is known as a Karman vortex street. It is also possible to see the boundary layer on the plate in the figure on the left. What is interesting is that the fluid used in this experiment is water, and not air. This is indicative of the fact that the separation of the flow into inviscid and boundary layer domains is a characteristic of any fluid governed by the Navier-Stokes equations, assuming the Reynolds number is sufficiently large. It is also evident, given the complexity of the flow, that finding the solution for the finite plate requires numerical methods. Some of the issues that arise with this are discussed in Cebeci and Cousteix (2005).

Before closing this section, a couple of comments are needed about the boundary layer reduction. First, the flow in the immediate vicinity of the leading edge requires a more refined boundary layer analysis that was used here. The same comment applies to the trailing edge for the finite length plate. Second, there are questions remaining about the matching requirement for the vertical velocity. In particular,



**Fig. 9.18** Wake behind a flat plate, showing the vortices generated in the flow (Taneda 1958). The photograph of the left is the flow immediately behind the plate, and the one on the right is further downstream. The vortices are evident because aluminum particles are suspended in the flow. In this experiment,  $Re = 15,800$

there must be a matching condition, yet it is not included in (9.62). This is an issue, because according to (9.71), it appears that the vertical velocity is unbounded when one moves out of the boundary layer into the outer region. Namely, given that  $f'(\infty) = 1$ , then  $\eta f'$  is unbounded as  $\eta \rightarrow \infty$ . In comparison, we know that the vertical velocity in the inviscid region is just zero. Therefore, to guarantee that the vertical velocity matches it must be that  $f \sim \eta$  as  $\eta \rightarrow \infty$ . If the solution of (9.68) does not do this, then the whole approximation fails. It is found, from the numerical solution, that  $f$  does indeed have the correct limiting behavior, indicating that the expansions match.

## 9.6 Water Waves

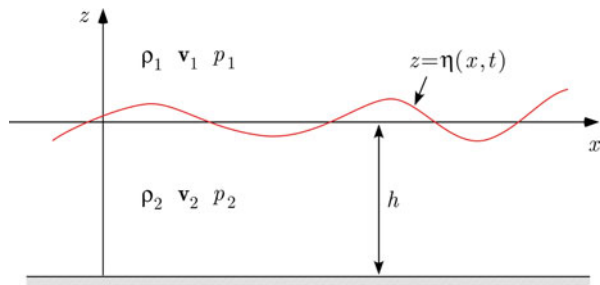
Some of the most interesting problems in fluid dynamics involve wave motion. The example everyone is familiar with is water waves, as occur on a lake or ocean. We are going to consider two problems associated with water waves. The first will involve the characteristic properties of the waves, and the second will examine a mechanism that is responsible for generating the waves. In both cases, it is assumed that the wave motion is two-dimensional, as illustrated in Fig. 9.19. The reasoning here is that in the  $y$ -direction, which is perpendicular to the direction the wave is moving, there appears to be little motion for a typical water wave.

As indicated in Fig. 9.19, there will be equations for the upper fluid (the air), and equations for the lower fluid (the water). It is assumed that both the air and water are irrotational and ideal fluids. In this case, in their respective regions (see Fig. 9.19), we need to solve

$$\nabla^2 \phi_j = 0, \quad (9.72)$$

where  $j = 1$  refers to the air, and  $j = 2$  is for the water. Once  $\phi_j$  is determined, then  $u_j = \partial_x \phi_j$  and  $w_j = \partial_z \phi_j$ . Also, the pressure  $p_j$  is determined using Bernoulli's theorem (9.29). Assuming gravity is the body force, then

**Fig. 9.19** Two fluids, in the  $x, z$ -plane, that are separated by an interface curve  $z = \eta(x, t)$



$$p_j = p_{0j}(t) - \rho_j \left( \frac{\partial \phi_j}{\partial t} + \frac{1}{2} \nabla \phi_j \cdot \nabla \phi_j + gz \right). \quad (9.73)$$

The water occupies  $-h < z < \eta$ , and assuming the bottom is solid and not moving, the boundary condition is

$$\frac{\partial \phi_2}{\partial z} = 0, \text{ for } z = -h. \quad (9.74)$$

The air occupies  $\eta < z < \infty$ , and it is assumed that

$$\frac{\partial \phi_1}{\partial z} = 0, \text{ for } z \rightarrow \infty. \quad (9.75)$$

The complication in this problem is determining the location of the interface curve  $z = \eta(x, t)$ . Because  $\eta(x, t)$  is one of the unknowns, this is an example of what is called a free-surface problem. Determining the appropriate interface conditions is necessary to continue.

### 9.6.1 Interface Conditions

The requirements we are looking for involve continuity of the velocity and stress across the interface. Because both fluids are inviscid, the specific requirements are that the normal velocity and pressure are continuous (see Sect. 8.11.2). Since the equation for the interface is  $z = \eta(x, t)$ , then a normal direction is  $\mathbf{n} = (\eta_x, -1)/\sqrt{\eta_x^2 + 1}$ . The resulting conditions are, when  $z = \eta(x, t)$ ,

$$\mathbf{v}_1 \cdot \mathbf{n} = \mathbf{v}_2 \cdot \mathbf{n}, \quad (9.76)$$

and

$$p_1(x, t) = p_2(x, t). \quad (9.77)$$

To write the normal velocity condition in component form, note that given a particle on the interface curve, its position can be written as  $\mathbf{r}(t) = (r_1(t), r_2(t))$ , and its velocity is  $\mathbf{r}'(t) = (r_1'(t), r_2'(t))$ . The continuity of normal velocity requirement (9.76) can now be written as

$$\mathbf{v}_j \cdot \mathbf{n} = \mathbf{r}' \cdot \mathbf{n}, \text{ for } j = 1, 2.$$

Since  $r_2 = \eta(r_1, t)$ , then  $r'_2 = (\partial_x \eta)r'_1 + \partial_t \eta$ . Substituting this into the above equation, and simplifying, the conclusion is that

$$w_j - u_j \frac{\partial \eta}{\partial x} = \frac{\partial \eta}{\partial t}, \text{ for } z = \eta(x, t), \quad (9.78)$$

where  $\mathbf{v}_j = (u_j, w_j)$ .

### 9.6.2 Traveling Waves

The first question we consider is, what sort of water waves are possible? Answering this is challenging because, even though the potential equation (9.72) is linear, the interface condition (9.78) is nonlinear. So, we will make two simplifying assumptions. The first is that the waves have a small amplitude. What this means is that we are going to assume that

$$\phi_j \sim \varepsilon \bar{\phi}_j(x, z, t), \quad (9.79)$$

$$p_j \sim p_0 - \rho_j g z + \varepsilon \bar{p}_j(x, z, t), \quad (9.80)$$

$$\eta \sim \varepsilon \bar{\eta}(x, t). \quad (9.81)$$

The underlying assumption here is that  $\varepsilon \ll 1$ . Note that to facilitate the derivation, the reduction is being carried out without first nondimensionalizing the problem.

The second assumption is that there are plane wave solutions, which means that  $\bar{\phi}_j = g_j(z)e^{i(kx - \omega t)}$ ,  $\bar{p}_j = q_j(z)e^{i(kx - \omega t)}$ , and  $\bar{\eta} = \bar{\eta}_0 e^{i(kx - \omega t)}$ , where  $k > 0$ . In this case, the potential equation (9.72) reduces to solving

$$\frac{d^2 g_j}{dz^2} = k^2 g_j. \quad (9.82)$$

The general solution of this is  $g_j = A_j e^{kz} + B_j e^{-kz}$ . With the boundary conditions (9.74) and (9.75) we have that

$$g_2 = A \cosh(k(z + h)), \quad (9.83)$$

and

$$g_1 = B e^{-kz}. \quad (9.84)$$

The  $O(\varepsilon)$  interface condition coming from (9.78) is (see Exercise 9.23)

$$\partial_z \bar{\phi}_j(x, 0, t) = \partial_t \bar{\eta}(x, t). \quad (9.85)$$

From this we get that  $A = -i\omega\bar{\eta}_0/(k \sinh kh)$  and  $B = i\omega\bar{\eta}_0/k$ . Finally, the  $O(\varepsilon)$  requirement coming from the pressure equation (9.77) is

$$\rho_1 \left( \partial_t \bar{\phi}_1(x, 0, t) + g\bar{\eta} \right) = \rho_2 \left( \partial_t \bar{\phi}_2(x, 0, t) + g\bar{\eta} \right). \quad (9.86)$$

Substituting our formulas for  $\bar{\phi}_1$ ,  $\bar{\phi}_2$ , and  $\bar{\eta}$  into the above equation, one obtains

$$\omega^2 = \frac{gk(\rho_2 - \rho_1)}{\rho_2 + \rho_1 \tanh kh} \tanh kh. \quad (9.87)$$

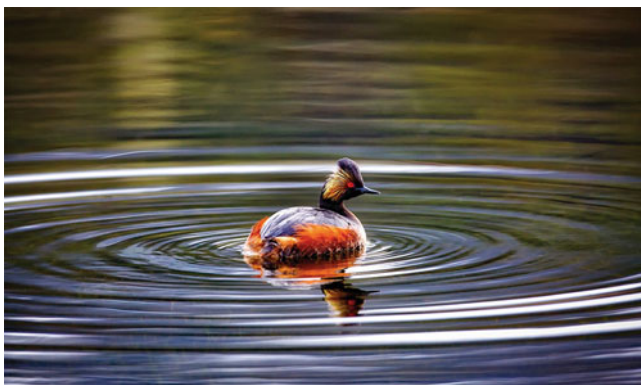
Therefore, if there is to be a traveling wave solution it must be that the frequency  $\omega$  and wavenumber  $k$  satisfy (9.87). Because the resulting frequency depends nonlinearly on the wavenumber, the problem is said to be dispersive.

### 9.6.2.1 Interpretation of Solution

To explore some of the consequences of (9.87), note that since  $\rho_1 \ll \rho_2$  and  $0 \leq \tanh kh \leq 1$ , then (9.87) can be approximated with

$$\omega^2 = gk \tanh kh. \quad (9.88)$$

In the case of deep water, which means that  $1 \ll kh$ , the wave frequency is  $\omega \approx \sqrt{gk}$ . Recall that for a plane wave  $e^{i(kx - \omega t)}$ , where  $\omega > 0$ , the phase velocity is  $v_{ph} = \omega/k$ . What we have determined is that for deep water  $v_{ph} = \sqrt{g/k}$ . Consequently, the longer the wave, the faster the wave moves. An illustration of this is given in Fig. 9.20.



**Fig. 9.20** Small amplitude waves (ripples in this case) created by a disturbance at the center, showing that the longer waves move faster than the shorter ones

In contrast, for shallow water, so  $kh \ll 1$ ,  $\omega \approx \sqrt{gh}k$ . The resulting phase velocity is  $v_{ph} \approx \sqrt{gh}$ , and so the velocity is independent of the wavelength. An interesting example of these waves is a tsunami. It can have wavelength as large as 500 km. In this case, since in the deep ocean the typical depth is 4 km,  $kh \approx 0.05$ . It also satisfies our small amplitude requirement, as in the deep ocean they have amplitudes of only about 1 m. This means a tsunami can be described using a shallow water approximation. What is impressive is that the resulting velocity is  $v_{ph} \approx 700$  kph (435 mph)! However, when it approaches a shoreline it slows down, and its amplitude grows, sometimes reaching up to 15 m. Because the amplitude is no longer small compared to the depth, the shallow water approximation is no longer valid. This requires an analysis of the nonlinear wave problem, and those interested in this should consult Zahibo et al. (2006) or Levin and Nosov (2016).

### 9.6.3 Wave Generation

What is of interest here is how waves are generated. Although there are a myriad of causes for waves, including boats and earthquakes, the principal cause is wind. To explore the mathematical problem for wind generated waves, we will consider the problem when two fluid layers are separated as shown in Fig. 9.19. As before, it's assumed that  $\rho_2 > \rho_1$ . The two fluids do not need to be water and air. For example, they could be layers of air at different temperatures, which occurs with an inversion layer, or two liquids with different densities.

To make any headway in solving this problem we need to identify the appropriate simplifying assumptions. The particular waves we will consider are shown in Fig. 9.21, and are due to what is called a Kelvin-Helmholtz instability. Two essential observations about the situation are that the waves can appear when the two fluids are air and/or water, and the velocities can be relatively low (as compared to the speed of sound). The subsequent assumptions, in respective order, are that the fluids can be assumed to be inviscid and incompressible. In other words, we will assume that both are ideal fluids. Also, from Fig. 9.21 it does not appear that the depths of the two fluid layers play an essential role. Consequently, it is assumed that the upper layer occupies  $\eta < z < \infty$ , and the lower layer occupies  $-\infty < z < \eta$ . Finally, it is assumed that the motion is in the  $x, z$ -plane, and the only body force is gravity.

#### 9.6.3.1 Derivation of Solution

Based on the stated assumptions, from (9.31) and (9.32), the equations for each layer are (for  $j = 1, 2$ )

$$\frac{\partial u_j}{\partial t} + u_j \frac{\partial u_j}{\partial x} + w_j \frac{\partial u_j}{\partial z} = -\frac{1}{\rho_j} \frac{\partial p_j}{\partial x}, \quad (9.89)$$



**Fig. 9.21** Examples of waves generated by a Kelvin-Helmholtz instability. Upper left: Laboratory experiment involving a tilt tank to generate waves (Thorpe, 1968). Lower left: Clouds formed at the interface of an inversion layer. Upper right: Event study involving the measurements of the interface between Earth's magnetosphere and the solar wind (Kavosi, 2015). Lower right: Cloud formation over Breckenridge, Colorado (Iskryan, 2019)

$$\frac{\partial w_j}{\partial t} + u_j \frac{\partial w_j}{\partial x} + w_j \frac{\partial w_j}{\partial z} = -\frac{1}{\rho_j} \frac{\partial p_j}{\partial z} - g, \quad (9.90)$$

$$\frac{\partial u_j}{\partial x} + \frac{\partial w_j}{\partial z} = 0, \quad (9.91)$$

where  $j = 1$  refers to the upper layer, and  $j = 2$  is for the lower layer. There are two conditions imposed at the interface. One is the continuity of the normal velocity (9.78), which is

$$w_j - u_j \frac{\partial \eta}{\partial x} = \frac{\partial \eta}{\partial t}, \quad \text{for } z = \eta(x, t). \quad (9.92)$$

The second condition is that the pressure is continuous, and so, from (9.77), it is required that  $p_1 = p_2$  when  $z = \eta(x, t)$ .

In preparation for generating waves, we will assume that the flow in each layer, at the start, is horizontal and there are no waves. Mathematically, the assumption is that  $\mathbf{v}_1 = (U_1, 0)$ ,  $\mathbf{v}_2 = (U_2, 0)$ ,  $p_1 = p_0 - \rho_1 g z$ ,  $p_2 = p_0 - \rho_2 g z$ , and the interface is simply  $z = 0$ . The velocities  $U_1$  and  $U_2$  are constants, as is  $p_0$ . It is easy to verify that these functions satisfy (9.89)–(9.91), as well as the stated conditions at the interface.

Our approach to studying the generation of waves uses a small disturbance approximation, similar to what was used in Sect. 5.6.1. The idea is that there is a



small disturbance in the flow, and the expansions for the dependent variables have the form

$$u_j \sim U_j + \varepsilon \bar{u}_j(x, z, t), \quad (9.93)$$

$$w_j \sim \varepsilon \bar{w}_j(x, z, t), \quad (9.94)$$

$$p_j \sim p_0 - \rho_j g z + \varepsilon \bar{p}_j(x, z, t), \quad (9.95)$$

$$\eta \sim \varepsilon \bar{\eta}(x, t). \quad (9.96)$$

The underlying assumption here is that  $\varepsilon \ll 1$ . Note that to facilitate the derivation, the reduction is being carried out without first nondimensionalizing the problem. Plugging the above expansions into (9.89)–(9.91), one finds that the  $O(\varepsilon)$  problem is (for  $j = 1, 2$ )

$$\frac{\partial \bar{u}_j}{\partial t} + U_j \frac{\partial \bar{u}_j}{\partial x} = -\frac{1}{\rho_j} \frac{\partial \bar{p}_j}{\partial x}, \quad (9.97)$$

$$\frac{\partial \bar{w}_j}{\partial t} + U_j \frac{\partial \bar{w}_j}{\partial x} = -\frac{1}{\rho_j} \frac{\partial \bar{p}_j}{\partial z}, \quad (9.98)$$

$$\frac{\partial \bar{u}_j}{\partial x} + \frac{\partial \bar{w}_j}{\partial z} = 0. \quad (9.99)$$

The  $O(\varepsilon)$  interface condition coming from (9.78) is (see Exercise 9.23)

$$\bar{w}_j(x, 0, t) - U_j \partial_x \bar{\eta} = \partial_t \bar{\eta} \quad \text{for } j = 1, 2. \quad (9.100)$$

and from (9.77) we obtain

$$(\rho_2 - \rho_1)g\bar{\eta} + \bar{p}_1(x, 0, t) = \bar{p}_2(x, 0, t). \quad (9.101)$$

Now comes the last assumption. Although it is possible to solve the  $O(\varepsilon)$  problem exactly, our goal is to understand the physical mechanism that is responsible for the generation of the waves. To that end, we make the simplifying assumption that the  $x$  and  $t$  dependence of the disturbance is a plane wave of the form  $e^{i(kx - \omega t)}$ , where  $k > 0$ . In other words, it is assumed that  $\bar{u}_j = g_j(z)e^{i(kx - \omega t)}$ ,  $\bar{w}_j = h_j(z)e^{i(kx - \omega t)}$ ,  $\bar{p}_j = q_j(z)e^{i(kx - \omega t)}$ , and  $\bar{\eta} = \bar{\eta}_0 e^{i(kx - \omega t)}$ . With this, (9.97) and (9.98) reduce to

$$g_j(z) = \frac{k}{\rho_j(\omega - kU_j)} q_j(z), \quad (9.102)$$

$$h_j(z) = -\frac{i}{\rho_j(\omega - kU_j)} q_j'(z), \quad (9.103)$$

and (9.99) reduces to  $q_j'' = k^2 q_j$ . Solving this, and requiring the solution to be bounded, it follows that  $q_1 = q_{10}e^{-kz}$ , and  $q_2 = q_{20}e^{kz}$ , where  $q_{10}$  and  $q_{20}$  are constants. From (9.100), one finds that  $q_{j0} = (-1)^j \bar{\eta}_0 \rho_j (\omega - kU_j)^2 / k$ . This leaves solving (9.101), which reduces to solving the quadratic equation  $a\omega^2 - 2b\omega + c = 0$ , where

$$a = \frac{1}{k}(\rho_1 + \rho_2), \quad (9.104)$$

$$b = \rho_2 U_2 + \rho_1 U_1, \quad (9.105)$$

$$c = k(\rho_2 U_2^2 + \rho_1 U_1^2) - (\rho_2 - \rho_1)g. \quad (9.106)$$

The solutions are  $\omega = (b \pm \sqrt{b^2 - ac})/a$ . What this means is that for there to be a disturbance of the form we have assumed, then to be consistent with the equations of motion, it is required that  $\omega$  be one of the solutions of the quadratic equation.

### 9.6.3.2 Interpretation of Solution

Our interest is the generation of waves, and so we concentrate on the solution for the interface  $z = \eta(x, t)$ . We have determined that if a small plane wave disturbance is introduced into the flow, then

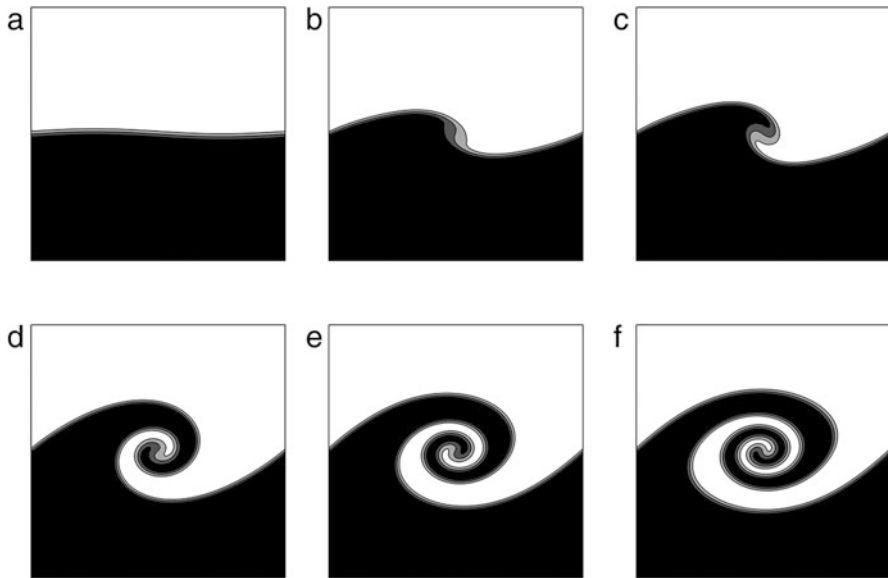
$$\eta \sim \eta_0 e^{i(kx - \omega t)}, \quad (9.107)$$

where  $\omega = (b \pm \sqrt{b^2 - ac})/a$ . The values of  $a$ ,  $b$ , and  $c$  are given in (9.104)–(9.106). The critical question related to the generation of waves is whether  $\omega$  is real-valued or complex-valued. If it is real-valued, then our solution (9.107) says that tiny little disturbances remain tiny, and so no waves are generated. If  $\omega$  is complex-valued, then (9.107) can grow exponentially in time. This possibility gives rise to what is called a *Kelvin-Helmholtz instability*. The implication is that this instability initiates the growth of the small disturbances into much larger waves.

To apply our solution to the waves on a lake, suppose the water is at rest, so  $U_2 = 0$ . In this case,  $b^2 - ac = -\rho_1 \rho_2 U_1^2 + g(\rho_2^2 - \rho_1^2)/k$ . Using cgs units, and round numbers, then  $\rho_1 = 1$ ,  $\rho_2 = 10^{-3}$ , and  $g = 1000$ . This means that for a wave with a wavelength of 1 m,

$$b^2 - ac \approx 10^{-3}(10^7 - U_1^2). \quad (9.108)$$

So, if the speed of the wind is low enough that  $U_1^2 < 10^7$ , then the  $\omega$  values are real, and no larger waves are generated. In contrast, for higher speed winds, so  $U_1^2 > 10^7$ , then there is a complex-valued solution, and the generation of larger waves is possible. The critical value, that separates the case of no waves from the case of when larger waves are generated, is a wind speed of  $U_1 \approx 36$  m/s (80 mph).



**Fig. 9.22** Computational solution of the fluid problem used to derive the Kelvin-Helmholtz instability at (a)  $t = 0$ , (b)  $t = 0.6$ , (c)  $t = 0.7$ , (d)  $t = 1$ , (e)  $t = 1.2$ , and (f)  $t = 1.4$  (Lee and Kim, 2015)

If you think that this speed is higher than what actually occurs, you are correct and this will be discussed below.

The small disturbance approximation used to obtain (9.107) is useful for determining how the waves are initiated. Once the waves start to grow, the physical assumptions we made are no longer valid. This raises the questions as to how well our model predicts the onset of wave generation, and what happens to the waves once they are created. These can be answered by solving the two-dimensional fluid equations numerically, and an example of what is obtained is shown in Fig. 9.22. The computed wave structure is very similar to the physical examples shown in Fig. 9.21, which indicates that the fluid model is consistent with the physical problem being studied. It is also found that the growth rate predicted by the small disturbance approximation matches well with the computed solution.

Given the large wind speeds predicted by the Kelvin-Helmholtz instability, it is evident that something essential is missing in the model to describe waves on a lake. Considerable effort has been invested to determine what this is, and this includes accounting for the fluid viscosity, surface tension, turbulent flow in the air, etc. Although its applicability to waves on a lake is perhaps questionable, it is known that the Kelvin-Helmholtz instability mechanism plays a role in a wide variety of wave motion problems. Examples include those illustrated in Fig. 9.21, as well as a Dirac fluid made up very fast moving electrons (Coelho et al., 2017), or waves of amplitudes up to 100 m in the ocean (van Haren et al., 2014). Those interested in

looking into more realistic models for waves on a lake, or an ocean, should consult Cavaleri et al. (2007) and Paquier et al. (2016). Also, those interested in a more rigorous derivation of the small disturbance approximation should consult Shtemler et al. (2008).

## Exercises

### Section 9.1

**9.1** As a modification of the plane Couette flow problem, suppose there are two incompressible fluids between the plates. One fluid occupies the region  $0 < y < h_0$ , and has density  $\rho_1$  and viscosity  $\mu_1$ . The second fluid occupies the region  $h_0 < y < h$  and has density  $\rho_2$  and viscosity  $\mu_2$ .

- (a) In plane Couette flow the velocity has the form  $\mathbf{v} = (u(y), 0, 0)$ . Also, at the interface, where  $y = h_0$ , the velocity and stress are assumed to be continuous. Use this to show that  $p$ ,  $u$ , and  $u'(y)$  are continuous at  $y = h_0$ .
- (b) Using the results from part (a), solve this plane Couette problem.

**9.2** Instead of Poiseuille flow in a circular cylinder, as considered in Sect. 9.1.2, suppose the cylinder has a cross-sectional shape  $S$ . For example,  $S$  is an ellipse or an equilateral triangle. In this problem Cartesian, rather than polar, coordinates are used.

- (a) What are the boundary conditions?
- (b) Which simplifying assumptions made for the circular cross-sectional case still hold, and which are no longer valid?
- (c) Show that  $p$  is the same as in the circular cylinder case, and that the axial fluid velocity  $w(x, y)$  satisfies an equation of the form  $(\partial_x^2 + \partial_y^2)w = \alpha$ , where  $w = 0$  on the boundary  $\partial S$ .
- (d) Suppose a series solution is sought of the form

$$w = a_{00} + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2 \\ + a_{30}x^3 + a_{21}x^2y + a_{12}xy^2 + a_{03}y^3 + \cdots$$

What conditions do the above  $a_{ij}$ 's satisfy so that  $w$  satisfies the differential equation in part (c)?

- (e) Suppose that  $S$  is symmetric in  $x$ , which means that if  $S$  is flipped (or, reflected) through the  $y$ -axis, that the domain does not change. In this case,  $w$  must be an even function of  $x$  (you do not need to prove this). What conclusion can you make about the coefficients  $a_{10}$ ,  $a_{11}$ ,  $a_{30}$ , and  $a_{21}$ ? What if  $S$  is also symmetric in  $y$ ?

- (f) Find the axial velocity  $w$  in the case of when  $S$  is an ellipse.
- (g) Find the axial velocity  $w$  in the case of when  $S$  is an equilateral triangle.

**9.3** Suppose that in the plane Couette flow problem in Sect. 9.1.1 that gravity is included. This means that a forcing function must be included in (9.2) of the form  $\mathbf{f} = (0, -g, 0)$ .

- (a) What assumptions about the solution used to derive (9.6) no longer apply? What assumptions should still be valid?
- (b) Find the velocity  $\mathbf{v}$  and stress  $\boldsymbol{\sigma}$  for this flow.

## Section 9.2

**9.4** This problem examines the vorticity for a linear flow, which means that  $\mathbf{v} = \mathbf{H}\mathbf{x} + \mathbf{h}$ , where the matrix  $\mathbf{H}$  and vector  $\mathbf{h}$  can depend on  $t$ . Other properties of linear flows were developed in Exercises 8.5 and 8.19.

- (a) Show that  $\boldsymbol{\omega} = (H_{32} - H_{23}, H_{13} - H_{31}, H_{21} - H_{12})$ . What is the vorticity when  $\mathbf{H}$  is symmetric? What is the vorticity when  $\mathbf{H}$  is skew-symmetric?
- (b) The equations for vortex motion are given in Sect. 9.2.1. Show that the only vortex with a smooth velocity and constant vorticity has  $v_\theta = \omega r$ , where  $\omega$  is the constant angular velocity of the rotating flow. In this case, show that  $\mathbf{h} = \mathbf{0}$  and

$$\mathbf{H} = \begin{pmatrix} 0 & -\omega & 0 \\ \omega & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

- (c) As given in (8.13), rigid body rotation about the  $z$ -axis corresponds to the case of when  $\mathbf{b} = \mathbf{0}$  and  $\mathbf{Q}$  is given in (8.14). The resulting formula for the velocity is  $\mathbf{v} = \mathbf{Q}'\mathbf{Q}^T\mathbf{x}$ . Show that the  $\mathbf{H}$  found in part (b) corresponds to rigid body rotation about the  $z$ -axis.

**9.5** For a Taylor vortex,  $v_r = v_z = 0$ , and

$$v_\theta = \frac{\alpha r}{t^2} \exp\left(-r^2/(4\nu t)\right),$$

where  $\nu = \mu/\rho$  is the kinematic viscosity.

- (a) Show that this satisfies the continuity equation (9.3).
- (b) What is the pressure?
- (c) What is the vorticity?

**9.6** For Burger's vortex,  $v_r = -\alpha r$ ,  $v_z = 2\alpha z$ , and

$$v_\theta = \frac{\beta}{r} \left( 1 - e^{-\alpha r^2/(2\nu)} \right),$$

where  $\nu = \mu/\rho$  is the kinematic viscosity.

- (a) Show that this satisfies the continuity equation (9.3).
- (b) What is the pressure?
- (c) What is the vorticity?

**9.7** Suppose that the velocity of an incompressible fluid is  $\mathbf{v} = \mathbf{v}_0 + \frac{1}{2}\boldsymbol{\Omega} \times \mathbf{x}$ , where  $\mathbf{v}_0$  and  $\boldsymbol{\Omega}$  are constant vectors. Consequently,  $\mathbf{v}$  consists of a constant velocity  $\mathbf{v}_0$  added to the velocity for circular motion in the plane perpendicular to  $\boldsymbol{\Omega}$ .

- (a) Show that  $\boldsymbol{\omega} = \boldsymbol{\Omega}$ .
- (b) The helicity density is defined as  $h = \boldsymbol{\omega} \cdot \mathbf{v}$ , and it gives rise to the invariant derived in Exercise 9.18(b). Using the result from part (a), show that  $h = \boldsymbol{\Omega} \cdot \mathbf{v}_0$ .
- (c) Assuming  $\boldsymbol{\Omega} = (0, 0, \Omega)$  and  $\mathbf{v}_0 = (0, 0, w_0)$ , find the pathlines and from this show that the flow is helical.
- (d) From the description of  $\mathbf{v}$  given above, one might think that it corresponds to rigid body motion. Show that is actually true.

**9.8** This problem develops some of the properties of the vorticity.

- (a) Show that  $\nabla \cdot \boldsymbol{\omega} = 0$ .
- (b) Writing  $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)^T$ , show that vorticity tensor, defined in (8.67), is

$$\mathbf{W} = \frac{1}{2} \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix}.$$

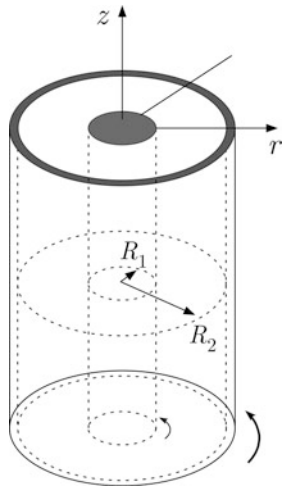
- (c) For two-dimensional flow, where  $\mathbf{v} = (u, v, 0)^T$ , show that  $\frac{D}{Dt}\boldsymbol{\omega} = \mathbf{0}$ . Consequently, two-dimensional flow that starts out irrotational, remains irrotational. This statement is not true for three-dimensional flow without additional assumptions (such as made in Helmholtz's Third Vorticity Theorem).
- (d) Show that for an incompressible fluid, with a conservative body force,

$$\frac{D}{Dt}\boldsymbol{\omega} = (\boldsymbol{\omega} \cdot \nabla)\mathbf{v} + \nu \nabla^2 \boldsymbol{\omega},$$

where  $\nu = \mu/\rho$  is the kinematic viscosity. This is known as the vorticity transport equation.

Hint:  $(\mathbf{v} \cdot \nabla)\mathbf{v} = \frac{1}{2}\nabla(\mathbf{v} \cdot \mathbf{v}) - \mathbf{v} \times \boldsymbol{\omega}$

**Fig. 9.23** Concentric rotating cylinders used in the Taylor-Couette problem in Exercise 9.9



**9.9** An incompressible viscous fluid occupies the region between two concentric cylinders of radii  $R_1$  and  $R_2$ , where  $R_1 < R_2$ . Assume the cylinders are infinitely long and centered on the  $z$ -axis (see Fig. 9.23). The inner cylinder is assumed rotating around the  $z$ -axis with angular velocity  $\omega_1$ , while the outer cylinder rotates around the  $z$ -axis with angular velocity  $\omega_2$ . The flow is assumed to be steady, and there are no body forces. This is known as the Taylor-Couette problem.

- Using cylindrical coordinates, explain why the boundary conditions on the cylinders are  $(v_r, v_\theta, v_z) = (0, \omega_i R_i, 0)$  for  $r = R_i$ .
- Explain why it is reasonable to assume that the solution has  $v_z = 0$  and  $v_r = 0$ .
- Find  $v_\theta$  and  $p$ .
- What is the vorticity for this flow? With this show that the flow is irrotational if  $R_1^2 \omega_1 = R_2^2 \omega_2$ .

**9.10** This exercise explores the connections between vorticity and energy dissipation in a viscous fluid.

- The viscous dissipation function  $\Phi$  is given in (8.109). Show that

$$\Phi = 2\mu(D_{xx}^2 + D_{yy}^2 + D_{zz}^2 + 2D_{xy}^2 + 2D_{xz}^2 + 2D_{yz}^2),$$

where the  $D_{ij}$ 's are the components of the rate of deformation tensor given in (8.66).

- Show that for an incompressible fluid,

$$\Phi = \mu \boldsymbol{\omega} \cdot \boldsymbol{\omega} + 2\mu \nabla \cdot \mathbf{q},$$

where  $\mathbf{q} = (\nabla \mathbf{v})\mathbf{v}$ .

- (c) Let  $B$  be a bounded region in space. Use the result from part (b) to derive what is known as the Bobileff-Forsyth formula, given as

$$\iiint_B \Phi \, dV = \mu \iiint_B \boldsymbol{\omega} \cdot \boldsymbol{\omega} \, dV + 2\mu \iint_{\partial B} \mathbf{n} \cdot \mathbf{q} \, dS.$$

- (d) If  $\mathbf{v} = \mathbf{0}$  on  $\partial B$  show that

$$\iiint_B \Phi \, dV = \mu \iiint_B \boldsymbol{\omega} \cdot \boldsymbol{\omega} \, dV.$$

This shows that the total energy dissipation in the region is determined by the magnitude of the vorticity vector.

- (e) Show that for an incompressible fluid, with no body force,

$$\frac{d}{dt} \iiint_{R(t)} \frac{1}{2} \rho \mathbf{v} \cdot \mathbf{v} \, dV = \iint_{\partial R(t)} \mathbf{g} \cdot \mathbf{n} \, dS - \mu \iiint_{R(t)} \boldsymbol{\omega} \cdot \boldsymbol{\omega} \, dV,$$

where  $\mathbf{g} = -p\mathbf{v} - \mu\boldsymbol{\omega} \times \mathbf{v} + 2\mu\mathbf{q}$ , and  $\mathbf{q}$  is given in part (b).

- (f) If the fluid is compressible, show that the generalization of the Bobileff-Forsyth formula is

$$\iiint_B \Phi \, dV = \iiint_B [(\lambda + 2\mu)\Theta^2 + \mu\boldsymbol{\omega} \cdot \boldsymbol{\omega}] \, dV + 2\mu \iint_{\partial B} \mathbf{n} \cdot \mathbf{q} \, dS,$$

where  $\Theta = \nabla \cdot \mathbf{v}$  and  $\mathbf{q} = (\nabla \mathbf{v})\mathbf{v} - \Theta\mathbf{v}$ .

## Section 9.3

**9.11** Suppose an incompressible viscous fluid has velocity  $\mathbf{v} = (u, v, 0)$ , with  $u = ax^2 + bxy + cy^2$ , where  $a, b$ , and  $c$  are constant.

- Find  $v$  assuming that  $v(x, 0, z) = 0$ .
- Find  $\sigma$ .
- For what values of  $a, b$ , and  $c$ , if any, is the flow irrotational?

**9.12** Suppose the velocity for an incompressible fluid is  $\mathbf{v} = (-\alpha y, \alpha x, \beta)$ , where  $\alpha$  and  $\beta$  are constants.

- Show that  $v$  satisfies the continuity equation.
- Assuming no external body forces, find the pressure.
- Is this flow rotational or irrotational?



- (d) For a fluid particle that starts at  $\mathbf{x} = \mathbf{x}_0$ , its position  $\mathbf{x} = \mathbf{X}(t)$  at later times is determined by solving  $\mathbf{X}'(t) = \mathbf{v}(\mathbf{X}, t)$ , where  $\mathbf{X}(0) = \mathbf{x}_0$ . The resulting curve is known as a pathline. Find the pathlines for the given velocity.
- (e) This is known as steady helical flow. Why?

**9.13** Suppose the velocity for an incompressible fluid is  $\mathbf{v} = (x + y, 3x - y, 0)$ .

- (a) Show that  $\mathbf{v}$  satisfies the continuity equation.
- (b) Assuming no external body forces, find the pressure.
- (c) Is this flow rotational or irrotational?
- (d) Find the pathlines (these are defined in Exercise 9.12(d)).
- (e) Use the result from part (d) to find the material description of the flow.

**9.14** Suppose that  $\mathbf{v} = \alpha \|\mathbf{x}\|^k \mathbf{x}$ , where  $k$  and  $\alpha$  are real numbers.

- (a) Show that the flow is irrotational.
- (b) Find a potential function  $\phi$  for this flow.
- (c) Show that this velocity function does not correspond to incompressible fluid motion, unless  $\alpha = 0$ .

**9.15** This problem derives the formulas for the potential functions in the Helmholtz Representation Theorem. Assume that  $D = \mathbb{R}^3$ .

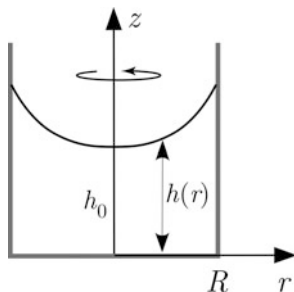
- (a) Given that  $\phi = \nabla \cdot \mathbf{h}$ , then from (9.22) derive (9.23).
- (b) Show that the vector potential function can be written as (9.24).

## Section 9.4

**9.16** An ideal fluid is rotating, with angular velocity  $\omega$ , inside a circular cylinder (i.e., a tin can) as illustrated in Fig. 9.24. As will be shown, assuming the body force is gravity, so  $\mathbf{f} = -g\mathbf{k}$ , then the fluid rotates as a rigid body. What is not known is the equation for the upper surface, and this is determined in the derivation.

- (a) Explain why the boundary condition for the fluid is that  $\mathbf{v} \cdot \mathbf{n} = 0$  on the sides and bottom of the can.

**Fig. 9.24** Rotating ideal fluid inside a tin can with radius  $R$



- (b) Show that the velocity for rigid body rotation given in Sect. 9.3 satisfies the boundary condition given in part (a).
- (c) Substitute the velocity for rigid body rotation into (9.31) and (9.32), and from this determine the general formula for the pressure.
- (d) On the upper surface the fluid pressure must equal the air pressure (which you can assume is just zero). What is the resulting equation  $z = h(r)$  for the upper surface? Your answer should include the constant  $h_0 = h(0)$ .
- (e) Explain what's wrong with Fig. 9.24 in the case of when the angular velocity is very large.
- (f) If the depth of the fluid, when at rest, is  $H$ , show that your answer in part (d) becomes

$$h(r) = H - \frac{1}{4g}\omega^2 R^2 \left[ 1 - 2\left(\frac{r}{R}\right)^2 \right].$$

Use this to explain why your solution in part (e) requires that  $\omega < 2\sqrt{gH}/R$ .  
 Comment: In fluid mechanics textbooks, the restriction of the angular velocity is typically written as  $Fr < 2$ , where  $Fr = \omega R/\sqrt{gH}$  is the Froude number for this flow.

**9.17** In this problem assume the body force in the Navier-Stokes equations can be written as  $\mathbf{f} = \nabla\psi$ .

- (a) Assuming that the fluid is ideal show that

$$\frac{\partial \mathbf{v}}{\partial t} + \nabla \left( \frac{1}{2} \mathbf{v} \cdot \mathbf{v} + \frac{1}{\rho} p - \psi \right) + \boldsymbol{\omega} \times \mathbf{v} = \mathbf{0}.$$

In the case where the fluid is also irrotational show that

$$p = p_0(t) - \rho \left( \frac{\partial \phi}{\partial t} + \frac{1}{2} \nabla \phi \cdot \nabla \phi \right) + \rho \psi.$$

- (b) Suppose the fluid is inviscid and irrotational. Also, assume it satisfies the equation of state for a polytropic fluid, which is  $p = k\rho^\gamma$ , where  $\gamma > 1$ . Adapt the argument of part (a) to show that

$$\frac{\partial \phi}{\partial t} + \frac{\gamma}{\gamma - 1} \frac{p}{\rho} + \frac{1}{2} \nabla \phi \cdot \nabla \phi - \psi = c(t).$$

**9.18** There are three known principal invariants, or conserved quantities, for an ideal fluid. One is the circulation, which comes from Kelvin's Circulation Theorem. This problem derives the other two. Assume  $R(t)$  is a material volume, as used for the Reynolds Transport Theorem. Also assume that the body force is conservative, so it can be written as  $\mathbf{f} = \nabla\psi$ .

- (a) If  $\mathbf{v} \cdot \mathbf{n} = 0$  on  $\partial R$ , show that

$$\frac{d}{dt} \iiint_R \mathbf{v} \cdot \mathbf{v} dV = 0.$$

This is the energy invariant, and states that the kinetic energy of the material volume is constant.

- (b) If  $\boldsymbol{\omega} \cdot \mathbf{n} = 0$  on  $\partial R$ , show that

$$\frac{d}{dt} \iiint_R \boldsymbol{\omega} \cdot \mathbf{v} dV = 0.$$

This is called the helicity invariant, and it measures the extent the pathlines coil around each other.

## Section 9.5

**9.19** For the impulsive plate problem in Sect. 9.5.1, suppose the lower plate moves with velocity  $\mathbf{v} = (u_0 f(t), 0, 0)$ . Assuming that  $f(t)$  is a smooth function of  $t$ , use the Laplace transform to show that

$$u(y, t) = u_0 f(0) \operatorname{erfc}\left(\frac{y}{2\sqrt{vt}}\right) + u_0 \int_0^t f'(t-r) \operatorname{erfc}\left(\frac{y}{2\sqrt{vr}}\right) dr.$$

**9.20** For the impulsive plate problem in Sect. 9.5.1, suppose the lower plate moves with velocity  $\mathbf{v} = (u_0 \cos(\omega t), 0, 0)$ . This is known as Stokes' second problem. The exact solution can be found using the formula from the previous problem. However, in this problem we are interested in the resulting periodic response, and for this reason we will derive the solution directly.

- After a sufficiently long period of time the solution should be approximately periodic. Assume that  $u = e^{i\omega t} q(y)$ , where it is understood that the real part of this expression is used. This expression should satisfy the momentum equation, and the two boundary conditions. Show that this results in a solution of the form  $u = u_0 e^{-\sigma y} \cos(\sigma y - \omega t)$ .
- Sketch the solution as a function of  $y$ , and describe the basic characteristics of the solution.
- Show that the boundary layer thickness is approximately  $5\sqrt{2\nu/\omega}$ . What happens to the thickness as the frequency increases?

## Section 9.6

**9.21** This problem explores how the unperturbed velocities  $U_1$  and  $U_2$  affect Kelvin-Helmholtz instability.

- Suppose both fluids are at rest, so  $U_1 = U_2 = 0$ . What is the conclusion coming from the stability analysis?
- Suppose  $U_1 = U_2 \neq 0$ . What is the conclusion coming from the stability analysis?
- Suppose  $U_1 = -U_2 \neq 0$ . What is the conclusion coming from the stability analysis?

**9.22** This problem explores some of the properties of the waves coming from the analysis of the Kelvin-Helmholtz instability. Assume that the lower fluid is initially at rest, so  $U_2 = 0$ . Also, assume that  $U_1 > 0$  and  $k > 0$ .

- Suppose that  $\omega = \omega_r + i\omega_i$ , where  $\omega_r$  and  $\omega_i$  are the real and imaginary parts of  $\omega$ , respectively. For a plane wave  $e^{i(kx - \omega t)}$ , the phase velocity is  $v_{ph} = \omega_r/k$ . For the stable case, so  $b^2 - ac > 0$ , are the waves obtained from (9.107) faster or slower than the upper fluid? Is it possible for the waves to go in the opposite direction than the upper fluid?
- Redo part (a) for the unstable case, so  $b^2 - ac < 0$ .
- Given values for the densities and velocities, show that there is a  $k_0$  so that any small plane-wave disturbance with  $k_0 < k$  can produce an unstable wave.

**9.23** This problem fills in some of the steps in the derivation of the linearization of the interface conditions.

- Suppose  $g(x, z, t, \varepsilon) = 0$  on the curve given as  $z = \eta(x, t)$ . Assuming that  $\eta \sim \varepsilon\eta_0(x, t) + \varepsilon^2\eta_1(x, t) + \dots$ . Show that the  $O(1)$  requirement is that  $g(x, 0, t, 0) = 0$ , and the  $O(\varepsilon)$  requirement is that  $\eta_0(x, t)\partial_z g(x, 0, t, 0) + \partial_\varepsilon g(x, 0, t, 0) = 0$ .
- Assuming (9.79) and (9.81) hold, use the results from part (a) to derive (9.85).
- Assuming (9.79) and (9.81) hold, use the results from part (a) to derive (9.86).

**9.24** This problem explores the effect on the Kelvin-Helmholtz instability when surface tension of the interface is included. The assumption is that surface tension introduces a stress that is proportional to the mean curvature of the interface curve. In fluid dynamics, the usual way this statement is expressed is

$$p_1 = p_2 + \sigma \frac{\partial}{\partial x} \left( \frac{\partial_x \eta}{\sqrt{1 + (\partial_x \eta)^2}} \right), \quad \text{when } z = \eta(x, t),$$

where  $\sigma$  is the surface tension (it is a positive constant). This condition replaces the requirement that  $p_1 = p_2$ , when  $z = \eta(x, t)$ .

- How does the inclusion of surface tension change (9.101)? Note that Exercises 9.23(b),(c) might be helpful here.

- (b) Show that  $\omega$  stills satisfies a quadratic equation, and the solutions are

$$\omega = k \frac{\rho_2 U_2 + \rho_1 U_1}{\rho_1 + \rho_2} \pm k \sqrt{\frac{g(\rho_2 - \rho_1) + k^2 \sigma}{k(\rho_1 + \rho_2)} - \rho_1 \rho_2 \left( \frac{U_2 - U_1}{\rho_1 + \rho_2} \right)^2}.$$

- (c) For the waves on a lake example considered in Sect. 9.6.3.2,  $U_2 = 0$ . Assuming this holds, explain why the inclusion of surface tension results in the critical value for  $U_1$ , that separates the case of no waves from the case of when larger waves are generated, to increase. What this means is that surface tension stabilizes the interface, resulting in the need for higher wind speeds to generate larger waves.
- (d) It is often stated that for very short waves, what are called capillary waves or ripples, that surface tension is the principal mechanism determining the critical value for  $U_1$ . In contrast, for longer waves, what are called gravity waves, that gravity is the principal mechanism determining the critical value for  $U_1$ . Explain the reasoning for these statements using the result from part (b).
- (e) Derive the formula in Exercise 1.6(b).

### Additional Questions

**9.25** This problem examines a model for power-law fluids. It is based on the observation coming from Fig. 9.3 that the shear stress for plane Couette flow has the form  $\sigma_{12} = \alpha \left( \frac{\partial u}{\partial y} \right)^\beta$ . It is assumed here that the fluid is incompressible.

- (a) In plane Couette flow the velocity has the form  $\mathbf{v} = (u(y), 0, 0)$ . What are  $\mathbf{D}$  and its three invariants in this case?
- (b) As shown in Sect. 8.10.2.1, the general form of the constitutive law for a nonlinear viscous fluid is  $\boldsymbol{\sigma} = -p\mathbf{I} + \mathbf{G}$ , where  $\mathbf{G} = \alpha_0\mathbf{I} + \alpha_1\mathbf{D} + \alpha_2\mathbf{D}^2$ . Explain how the power-law

$$\sigma_{12} = \alpha \left| \frac{\partial u}{\partial y} \right|^m \frac{\partial u}{\partial y}$$

is obtained by assuming that  $\alpha_0 = \alpha_2 = 0$  and  $\alpha_1$  depends on  $\Pi_D$  in a particular way. It is assumed that  $m > -1$ , which guarantees that  $\mathbf{G} = \mathbf{0}$  if  $\mathbf{D} = \mathbf{0}$ .

- (c) Assuming that  $\frac{\partial u}{\partial y} > 0$ , and using the constitutive law from part (b), solve the resulting plane Couette flow problem. From this show that  $\sigma_{12} = \alpha \dot{\gamma}^{m+1}$ , where  $\dot{\gamma}$  is given in (9.7).
- (d) On the same axes, sketch  $\sigma_{12}$  as function of  $\dot{\gamma}$  when  $-1 < m < 0$ , when  $m = 0$ , and when  $1 < m$ . Use this to compare the differences in the behavior of the shear stress for large values of  $\dot{\gamma}$ . Would the  $-1 < m < 0$  case be called a shear-thickening or a shear-thinning situation?

# Appendix A

## Taylor's Theorem

### A.1 Single Variable

The single most important result needed to develop an asymptotic approximation is Taylor's theorem. The single variable version of the theorem is below.

**Theorem A.1** *Given a function  $f(x)$ , assume that its  $(n+1)$ st derivative  $f^{(n+1)}(x)$  is continuous for  $x_L < x < x_R$ . In this case, if  $a$  and  $x$  are points in the interval  $(x_L, x_R)$ , then*

$$f(x) = f(a) + (x-a)f'(a) + \frac{1}{2}(x-a)^2 f''(a) + \cdots + \frac{1}{n!}(x-a)^n f^{(n)}(a) + R_{n+1}, \quad (\text{A.1})$$

where the remainder is

$$R_{n+1} = \frac{1}{(n+1)!}(x-a)^{n+1} f^{(n+1)}(\eta), \quad (\text{A.2})$$

and  $\eta$  is a point between  $a$  and  $x$ .

There are different, but equivalent, ways to write the above result. One is

$$f(x+h) = f(x) + hf'(x) + \frac{1}{2}h^2 f''(x) + \cdots + \frac{1}{n!}h^n f^{(n)}(x) + R_{n+1}. \quad (\text{A.3})$$

The requirement here is that  $x$  and  $x+h$  are points in the interval  $(x_L, x_R)$ .

### A.1.1 Simplification via Substitution

Perturbation expansions are used extensively in the text, and they are usually found using Taylor's theorem. How much work this requires can often be reduced by using some of the basic properties of a Taylor series. Of particular importance is the substitution rule. The following examples illustrate how substitution can be used in conjunction with some of the relatively simple Taylor series expansions given in Table 2.1.

#### Examples

1. Find the Taylor series, about  $x = 0$ , of  $\frac{1}{1-x}$ .

Setting  $y = -x$ , then  $y = 0$  is equivalent to  $x = 0$ . With this,

$$\begin{aligned}\frac{1}{1-x} &= \frac{1}{1+y} \\ &= 1 - y + y^2 - y^3 + \cdots \\ &= 1 + x + x^2 + x^3 + \cdots \quad \blacksquare\end{aligned}$$

2. Find the Taylor series, about  $x = 0$ , of  $\cos(3x)$ .

Setting  $y = 3x$ , then  $y = 0$  is equivalent to  $x = 0$ . With this,

$$\begin{aligned}\cos(3x) &= \cos y \\ &= 1 - \frac{1}{2}y^2 + \frac{1}{4!}y^4 + \cdots \\ &= 1 - \frac{1}{2}(3x)^2 + \frac{1}{4!}(3x)^4 + \cdots \quad \blacksquare\end{aligned}$$

3. Find the Taylor series, about  $x = 0$ , of  $\sqrt{1+x^2}$ .

Setting  $y = x^2$ , then  $y = 0$  is equivalent to  $x = 0$ . With this,

$$\begin{aligned}\sqrt{1+x^2} &= \sqrt{1+y} \\ &= 1 + \frac{1}{2}y - \frac{1}{8}y^2 + \cdots \\ &= 1 + \frac{1}{2}x^2 - \frac{1}{8}x^4 + \cdots \quad \blacksquare\end{aligned}$$

To use substitution with Taylor's theorem, you are mostly limited to using integer powers of  $x$ . This is not the case when using substitution for obtaining an asymptotic expansion. To illustrate how this is done, variations of the above examples are used for finding an asymptotic expansion.

*Examples*

1. Find a four term expansion of  $\frac{1}{1-\sqrt{\varepsilon}}$ .

Setting  $y = -\sqrt{\varepsilon}$ , then  $y = 0$  is equivalent to  $\varepsilon = 0$ . With this,

$$\begin{aligned}\frac{1}{1-\sqrt{\varepsilon}} &= \frac{1}{1+y} \\ &= 1 - y + y^2 - y^3 + \cdots \\ &= 1 + \sqrt{\varepsilon} + \varepsilon + \varepsilon^{3/2} + \cdots \quad \blacksquare\end{aligned}$$

2. Find a three term expansion of  $\cos(3(\varepsilon + \varepsilon^2))$

Setting  $y = 3(\varepsilon + \varepsilon^2)$ , then  $y = 0$  when  $\varepsilon = 0$ . With this,

$$\begin{aligned}\cos(3(\varepsilon + \varepsilon^2)) &= \cos y \\ &= 1 - \frac{1}{2}y^2 + \frac{1}{4!}y^4 + \cdots \\ &= 1 - \frac{9}{2}(\varepsilon^2 + 2\varepsilon^3 + \varepsilon^4) + \frac{27}{8}(\varepsilon^4 + \cdots) + \cdots \\ &= 1 - \frac{9}{2}\varepsilon^2 - 9\varepsilon^3 + \cdots \quad \blacksquare\end{aligned}$$

3. Find a three term expansion of  $\sqrt{1 + (\sin \varepsilon)^2}$ .

Setting  $y = (\sin \varepsilon)^2$ , then  $\varepsilon = 0$  means that  $y = 0$ . With this,

$$\begin{aligned}\sqrt{1 + (\sin \varepsilon)^2} &= \sqrt{1 + y} \\ &= 1 + \frac{1}{2}y - \frac{1}{8}y^2 + \cdots \\ &= 1 + \frac{1}{2}(\sin \varepsilon)^2 - \frac{1}{8}(\sin \varepsilon)^4 + \cdots \\ &= 1 + \frac{1}{2}\left(\varepsilon - \frac{1}{6}\varepsilon^3 + \cdots\right)^2 - \frac{1}{8}\left(\varepsilon - \frac{1}{6}\varepsilon^3 + \cdots\right)^4 + \cdots \\ &= 1 + \frac{1}{2}\varepsilon^2 - \frac{7}{24}\varepsilon^4 + \cdots \quad \blacksquare\end{aligned}$$

**A.2 Two Variables**

The two-variable version of the expansion in (A.3) is

$$f(x+h, t+k) = f(x, t) + Df(x, t) + \frac{1}{2}D^2 f(x, t) + \cdots + \frac{1}{n!}D^n f(x, t) + R_{n+1}, \quad (\text{A.4})$$



where

$$D = h \frac{\partial}{\partial x} + k \frac{\partial}{\partial t}.$$

Writing this out, through quadratic terms, yields

$$\begin{aligned} f(x+h, t+k) &= f(x, t) + hf_x(x, t) + kf_t(x, t) \\ &\quad + \frac{1}{2}h^2 f_{xx}(x, t) + hk f_{xt}(x, t) + \frac{1}{2}k^2 f_{tt}(x, t) + \cdots. \end{aligned}$$

The subscripts in the above expression denote partial differentiation. So, for example,

$$f_{xt} = \frac{\partial^2 f}{\partial x \partial t}.$$

It is assumed that the function  $f$  has continuous partial derivatives up through order  $n+1$ .

The above expansion can be expressed in a form similar to the one in (A.1), and the result is

$$\begin{aligned} f(x, t) &= f(a, b) + (x-a)f_x(a, b) + (t-b)f_t(a, b) \\ &\quad + \frac{1}{2}(x-a)^2 f_{xx}(a, b) + (x-a)(t-b)f_{xt}(a, b) + \frac{1}{2}(t-b)^2 f_{tt}(a, b) \\ &\quad + \cdots. \end{aligned}$$

### A.3 Multivariable Versions

For more than two variables it is convenient to use vector notation. In this case (A.4) takes the form

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + Df(\mathbf{x}) + \frac{1}{2}D^2 f(\mathbf{x}) + \cdots + \frac{1}{n!}D^n f(\mathbf{x}) + R_{n+1},$$

where  $\mathbf{x} = (x_1, x_2, \dots, x_k)$ ,  $\mathbf{h} = (h_1, h_2, \dots, h_k)$  and

$$\begin{aligned} D &= \mathbf{h} \cdot \nabla \\ &= h_1 \frac{\partial}{\partial x_1} + h_2 \frac{\partial}{\partial x_2} + \cdots + h_k \frac{\partial}{\partial x_k}. \end{aligned}$$

Writing this out, through quadratic terms, yields

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \mathbf{h} \cdot \nabla f(\mathbf{x}) + \frac{1}{2} \mathbf{h}^T \mathbf{H} \mathbf{h} + \cdots,$$

where  $\mathbf{H}$  is the Hessian and is given as

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_k \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_k \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_k} & \frac{\partial^2 f}{\partial x_2 \partial x_k} & \cdots & \frac{\partial^2 f}{\partial x_k^2} \end{pmatrix}.$$

Taylor's theorem can also be extended to vector functions, although the formulas are more involved. To write down the expansion through the linear terms, assume that  $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x}))$ , and  $\mathbf{x} = (x_1, x_2, \dots, x_k)$ . In this case,

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) = \mathbf{f}(\mathbf{x}) + (\nabla \mathbf{f}) \mathbf{h} + \cdots,$$

where

$$\nabla \mathbf{f} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_k} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_k} \end{pmatrix}.$$

## Appendix B

### Fourier Analysis

#### B.1 Fourier Series

It is assumed here that the function  $f(x)$  is piecewise continuous for  $0 \leq x \leq \ell$ . Recall that this means  $f(x)$  is continuous on the interval  $0 \leq x \leq \ell$  except at a finite number of points within the interval at which the function has a jump discontinuity.

The Fourier sine series for  $f(x)$  is defined as

$$S(x) = \sum_{n=1}^{\infty} \beta_n \sin(\lambda_n x), \quad (\text{B.1})$$

where  $\lambda_n = n\pi/\ell$  and

$$\beta_n = \frac{2}{\ell} \int_0^{\ell} f(x) \sin(\lambda_n x) dx. \quad (\text{B.2})$$

The Fourier cosine series for  $f(x)$  is defined as

$$C(x) = \frac{1}{2} \alpha_0 + \sum_{n=1}^{\infty} \alpha_n \cos(\lambda_n x), \quad (\text{B.3})$$

where

$$\alpha_n = \frac{2}{\ell} \int_0^{\ell} f(x) \cos(\lambda_n x) dx. \quad (\text{B.4})$$

A certain amount of smoothness is required of the function  $f(x)$ , so the above series is defined. For example,  $f(x)$  must be smooth enough that the integrals in (B.2) and (B.4) exist. Certainly assuming  $f(x)$  is continuous is enough for the integrals,

but, unfortunately, this is not enough to guarantee that the series in (B.1) and (B.3) converge. They will converge, however, if  $f(x)$  and  $f'(x)$  are piecewise continuous. The question naturally arises as to what they converge to, and for this we have the following result.

**Theorem B.1** Assume  $f(x)$  and  $f'(x)$  are piecewise continuous for  $0 \leq x \leq \ell$ . On the interval  $0 < x < \ell$ , the Fourier sine series and the Fourier cosine series converge to  $f(x)$  at points where the function is continuous, and they converge to  $\frac{1}{2}(f(x+) + f(x-))$  at points where the function has a jump discontinuity. At the endpoints,  $S(0) = S(\ell) = 0$ , while  $C(0) = f(0^+)$  and  $C(\ell) = f(\ell^-)$ .

When using a Fourier series to solve a differential equation one usually needs the expansion of the solution as well as its derivatives. The problem is that it is not always possible to obtain the series for  $f'(x)$  by differentiating the series for  $f(x)$ . For example, given a sine series as in (B.1) one might be tempted to conclude that

$$S'(x) = \sum_{n=1}^{\infty} \beta_n \lambda_n \cos(\lambda_n x).$$

The issue is that the differentiation has resulted in  $\lambda_n$  appearing in the coefficient. As an example, for the function

$$f(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 2 & \text{if } 1 < x \leq 2, \end{cases}$$

one finds that

$$\beta_n \lambda_n = \frac{2}{\ell} [1 - 2(-1)^n + \cos(n\pi/2)].$$

The general term  $\beta_n \lambda_n \cos(\lambda_n x)$  of the series does not converge to zero as  $n \rightarrow \infty$ , and this means that the series does not converge. Consequently, additional restrictions must be imposed on  $f(x)$  to guarantee convergence. Basically what are needed are conditions that will give us  $\beta_n = O(1/n^2)$ , and this brings us to the next result.

**Theorem B.2** Assume  $f(x)$  is continuous, with  $f'(x)$  and  $f''(x)$  piecewise continuous, for  $0 \leq x \leq \ell$ . If  $f(x)$  is expanded in a cosine series, then the series for  $f'(x)$  can be found by differentiating the series for  $f(x)$ . If  $f(x)$  is expanded in a sine series, and if  $f(0) = f(\ell) = 0$ , then the series for  $f'(x)$  can be found by differentiating the series for  $f(x)$ .

The question of convergence for integration is much easier to answer. As long as the Fourier series of  $f(x)$  converges, then the series for the integral of  $f$  can be found by simply integrating the series for  $f$ .

## B.2 Fourier Transform

To derive the formula for the Fourier transform from the Fourier series, it is convenient to use the symmetric interval  $-\ell < x < \ell$ . Generalizing (B.2) and (B.3), the Fourier series of a continuous function  $f(x)$  is

$$f(x) = \frac{1}{2}\alpha_0 + \sum_{n=1}^{\infty} [\alpha_n \cos(\lambda_n x) + \beta_n \sin(\lambda_n x)],$$

where  $\lambda_n = n\pi/\ell$ ,

$$\alpha_n = \frac{1}{\ell} \int_{-\ell}^{\ell} f(x) \cos(\lambda_n x) dx,$$

and

$$\beta_n = \frac{1}{\ell} \int_{-\ell}^{\ell} f(x) \sin(\lambda_n x) dx.$$

By using the identities  $\cos(\theta) = \frac{1}{2}(e^{i\theta} + e^{-i\theta})$  and  $\sin(\theta) = \frac{1}{2i}(e^{i\theta} - e^{-i\theta})$ , the Fourier series can be written in exponential form as

$$f(x) = \sum_{n=-\infty}^{\infty} \gamma_n e^{i\lambda_n x},$$

where

$$\gamma_n = \frac{1}{2\ell} \int_{-\ell}^{\ell} f(\bar{x}) e^{-i\lambda_n \bar{x}} d\bar{x}.$$

Combining these two expressions

$$f(x) = \sum_{n=-\infty}^{\infty} \frac{1}{2\ell} \int_{-\ell}^{\ell} f(\bar{x}) e^{i\lambda_n(x-\bar{x})} d\bar{x}.$$

The sum in the above equation is reminiscent of the Riemann sum used to define integration. To make this more evident, let  $\Delta\lambda = \lambda_{n+1} - \lambda_n = \frac{\pi}{\ell}$ . With this

$$f(x) = \sum_{n=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\ell}^{\ell} f(\bar{x}) e^{i\lambda_n(x-\bar{x})} d\bar{x} \Delta\lambda.$$

The argument originally used by Fourier is that in the limit of  $\ell \rightarrow \infty$ , the above expression yields

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\bar{x}) e^{i\lambda(x-\bar{x})} d\bar{x} d\lambda.$$

Fourier then made the observation that the above equation can be written as  $f(x) = \mathcal{F}^{-1}(\mathcal{F}(f))$ , where  $\mathcal{F}$  is the Fourier transform defined in Sect. 4.5. With this, the Fourier transform was born.

To say that the above derivation is heuristic would be more than generous. However, it is historically correct, and it does show the origin of the Fourier transform and its inverse. The formal proof of the derivation can be found in Weinberger (1995).

## Appendix C

# Stochastic Differential Equations

The steps used to solve the Langevin equation look routine, and the solutions in (4.89) and (4.90) are not particularly remarkable. However, on closer inspection, the randomness of the forcing function raises some serious mathematical questions. An example of  $\mathbf{R}$  is shown in Fig. 4.27 using 400 points along the  $t$ -axis. As discussed in Sect. 4.8.1, the value of  $\mathbf{R}(t_1)$  is independent of the value of  $\mathbf{R}(t_2)$  if  $t_1 \neq t_2$ . This means that if more than 400 points are used, the graph would appear even more random than in Fig. 4.27. The question that immediately arises is whether the non-differentiability of this function causes the differential equation (4.88), or its solution (4.88), to be meaningless. One approach for addressing this issue rests on denial, where the calculations are carried out as if everything is just fine. This is, in fact, what was done to derive (4.89), and this approach almost works. To have it succeed, all that is needed is to make sense of the solution, and then use this to justify the entire process.

The question is, therefore, how to define the integrals in (4.89) and (4.90). The exponentials are not an issue, and so to simplify the discussion we will concentrate on the expression

$$\mathbf{W}(t) = \int_0^t \mathbf{R}(\tau) d\tau. \quad (\text{C.1})$$

The definition of this integral employs the same Riemann sum used in Calculus. With this in mind, we introduce a partition  $0 < t_1 < t_2 < \cdots < t_m < t$ , where  $t_0 = 0$  and  $t_{m+1} = t$ . For simplicity, it is assumed the points are equally spaced, and so  $t_{j+1} - t_j = \Delta t$ . Letting  $s_j$  be a point from the interval  $[t_j, t_{j+1}]$ , then we introduce the partial sum

$$\mathbf{S}_m = \sum_{j=0}^{m-1} \mathbf{R}(s_j) \Delta t. \quad (\text{C.2})$$

The question is, if  $\Delta t \rightarrow 0$ , does  $\mathbf{S}_m$  converge? The answer is yes, although convergence is measured in the mean-square sense. Knowing that it converges, then the limit of  $\mathbf{S}_m$  serves as the definition of the integral in (C.1). This definition preserves most, but not all, of the properties associated with standard integration. In particular,  $\mathbf{W}$  is a continuous function of  $t$ , and the integral is additive in the sense that if  $t_1 < t_2$ , then

$$\int_0^{t_2} \mathbf{R}(\tau) d\tau = \int_0^{t_1} \mathbf{R}(\tau) d\tau + \int_{t_1}^{t_2} \mathbf{R}(\tau) d\tau.$$

Moreover, the partial sums in (C.2) provide a method for numerically evaluating the stochastic integrals in (4.89) and (4.90).

Now that integration has been put onto a solid mathematical footing, we turn to the differential equation (4.88). In the case of when  $\mathbf{R}$  is smooth, this equation can be integrated to yield

$$\mathbf{v}(t) = \mathbf{v}(0) - \lambda \int_0^t \mathbf{v}(\tau) d\tau + \frac{1}{m} \int_0^t \mathbf{R}(\tau) d\tau. \quad (\text{C.3})$$

For smooth functions this integral equation is equivalent to the differential equation (4.88). This fact is used to explain what happens when a random forcing is used. Specifically, the interpretation of the differential equation (4.88) is that  $\mathbf{v}$  satisfies (C.3). It is for this reason that in the subject of stochastic differential equations, (4.88) is conventionally written using differentials as

$$d\mathbf{v} = -\lambda \mathbf{v} dt + \frac{1}{m} \mathbf{R} dt.$$

The implication in using this notation is that the stochastic differential equation is being interpreted as the solution of the associated integral equation. With this viewpoint, (C.1) can be written as  $d\mathbf{W} = \mathbf{R} dt$ . Those interested in pursuing the theoretical foundation of the stochastic differential equations should consult Oksendal (2003).



## Appendix D

### Identities

#### D.1 Trace

In the following,  $\mathbf{A}$  and  $\mathbf{B}$  are  $n \times n$  matrices, and  $\alpha$  and  $\beta$  are scalars.

$$\text{tr}(\alpha\mathbf{A} + \beta\mathbf{B}) = \alpha \text{tr}(\mathbf{A}) + \beta \text{tr}(\mathbf{B})$$

$$\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$$

$$\text{tr}(\mathbf{A}^T) = \text{tr}(\mathbf{A})$$

If  $\mathbf{A}$  is symmetric and  $\mathbf{B}$  is skew-symmetric, then  $\text{tr}(\mathbf{AB}) = 0$ .

#### D.2 Determinant

In the following,  $\mathbf{A}$  and  $\mathbf{B}$  are  $n \times n$  matrices, and  $\alpha$  and  $\beta$  are scalars.

$$\det(\mathbf{AB}) = \det(\mathbf{BA}) = \det(\mathbf{A})\det(\mathbf{B})$$

$$\det(\alpha\mathbf{A}) = \alpha^n \det(\mathbf{A})$$

$$\det(\mathbf{A}^T) = \det(\mathbf{A})$$

$$\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A})$$

$$\det(\mathbf{I}) = 1$$

### D.3 Vector Calculus

In the following,  $\phi(\mathbf{x})$  is a scalar-valued function,  $\mathbf{u}(\mathbf{x})$ ,  $\mathbf{v}(\mathbf{x})$  are vector-valued functions, and  $\mathbf{A}(\mathbf{x})$ ,  $\mathbf{B}(\mathbf{x})$  are  $3 \times 3$  matrix-valued functions of  $\mathbf{x}$ .

$$\nabla \cdot \mathbf{u} = \text{tr}(\nabla \mathbf{u})$$

$$\nabla \cdot (\phi \mathbf{u}) = \mathbf{u} \cdot \nabla \phi + \phi (\nabla \cdot \mathbf{u})$$

$$\nabla \cdot (\mathbf{A} \mathbf{u}) = \mathbf{u} \cdot (\nabla \cdot \mathbf{A}) + \text{tr}(\mathbf{A}^T \nabla \mathbf{u})$$

$$\nabla \cdot (\phi \mathbf{A}) = \mathbf{A} \nabla \phi + \phi (\nabla \cdot \mathbf{A})$$

$$\nabla \times (\nabla \phi) = \mathbf{0}$$

$$\nabla \cdot (\nabla \times \mathbf{u}) = 0$$

$$(\mathbf{u} \cdot \nabla) \mathbf{u} = \frac{1}{2} \nabla (\mathbf{u} \cdot \mathbf{u}) - \mathbf{u} \times (\nabla \times \mathbf{u})$$

$$\nabla \times (\nabla \times \mathbf{u}) = \nabla (\nabla \cdot \mathbf{u}) - \nabla^2 \mathbf{u}$$

$$(\mathbf{v} \cdot \nabla) \mathbf{u} = (\nabla \mathbf{u}) \mathbf{v}$$

$$(\nabla \mathbf{u}) \mathbf{v} + (\nabla \mathbf{v}) \mathbf{u} = \nabla (\mathbf{u} \cdot \mathbf{v})$$

In the above identities, letting  $\mathbf{u}(\mathbf{x}) = (u, v, w)$  and  $\mathbf{x} = (x, y, z)$ ,

$$\nabla \mathbf{u} = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} & \frac{\partial v}{\partial z} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} & \frac{\partial w}{\partial z} \end{pmatrix},$$

and

$$\nabla \cdot \mathbf{A} = \begin{pmatrix} \frac{\partial A_{11}}{\partial x} + \frac{\partial A_{12}}{\partial y} + \frac{\partial A_{13}}{\partial z} \\ \frac{\partial A_{21}}{\partial x} + \frac{\partial A_{22}}{\partial y} + \frac{\partial A_{23}}{\partial z} \\ \frac{\partial A_{31}}{\partial x} + \frac{\partial A_{32}}{\partial y} + \frac{\partial A_{33}}{\partial z} \end{pmatrix}.$$

### D.4 Miscellaneous

$$\mathbf{A}^T (\mathbf{A} \mathbf{u} \times \mathbf{A} \mathbf{v}) = \det(\mathbf{A}) (\mathbf{u} \times \mathbf{v})$$

# Appendix E

## Equations for a Newtonian Fluid

### E.1 Cartesian Coordinates

Letting  $\mathbf{v} = (u, v, w)$  and  $\mathbf{f} = (f, g, h)$ , then for an incompressible Newtonian fluid in Cartesian coordinates:

$$\begin{aligned}\rho \left( \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + w \frac{\partial u}{\partial z} \right) &= -\frac{\partial p}{\partial x} + \mu \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) + \rho f \\ \rho \left( \frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + w \frac{\partial v}{\partial z} \right) &= -\frac{\partial p}{\partial y} + \mu \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} \right) + \rho g \\ \rho \left( \frac{\partial w}{\partial t} + u \frac{\partial w}{\partial x} + v \frac{\partial w}{\partial y} + w \frac{\partial w}{\partial z} \right) &= -\frac{\partial p}{\partial z} + \mu \left( \frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2} \right) + \rho h \\ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} &= 0\end{aligned}$$

### E.2 Cylindrical Coordinates

Letting  $\mathbf{v} = (v_r, v_\theta, v_z) = v_r \mathbf{e}_r + v_\theta \mathbf{e}_\theta + v_z \mathbf{e}_z$  and  $\mathbf{f} = f_r \mathbf{e}_r + f_\theta \mathbf{e}_\theta + f_z \mathbf{e}_z$ , then for an incompressible Newtonian fluid in cylindrical coordinates:

$$\begin{aligned}\rho \left( \frac{\partial v_r}{\partial t} + v_r \frac{\partial v_r}{\partial r} + \frac{v_\theta}{r} \frac{\partial v_r}{\partial \theta} - \frac{v_\theta^2}{r} + v_z \frac{\partial v_r}{\partial z} \right) \\ = -\frac{\partial p}{\partial r} + \mu \left[ \frac{\partial}{\partial r} \left( \frac{1}{r} \frac{\partial}{\partial r} (r v_r) \right) + \frac{1}{r^2} \frac{\partial^2 v_r}{\partial \theta^2} - \frac{2}{r^2} \frac{\partial v_\theta}{\partial \theta} + \frac{\partial^2 v_r}{\partial z^2} \right] + \rho f_r\end{aligned}$$

$$\begin{aligned} \rho \left( \frac{\partial v_\theta}{\partial t} + v_r \frac{\partial v_\theta}{\partial r} + \frac{v_\theta}{r} \frac{\partial v_\theta}{\partial \theta} + \frac{v_r v_\theta}{r} + v_z \frac{\partial v_\theta}{\partial z} \right) \\ = -\frac{1}{r} \frac{\partial p}{\partial \theta} + \mu \left[ \frac{\partial}{\partial r} \left( \frac{1}{r} \frac{\partial}{\partial r} (r v_\theta) \right) + \frac{1}{r^2} \frac{\partial^2 v_\theta}{\partial \theta^2} + \frac{2}{r^2} \frac{\partial v_r}{\partial \theta} + \frac{\partial^2 v_\theta}{\partial z^2} \right] + \rho f_\theta \end{aligned}$$

$$\begin{aligned} \rho \left( \frac{\partial v_z}{\partial t} + v_r \frac{\partial v_z}{\partial r} + \frac{v_\theta}{r} \frac{\partial v_z}{\partial \theta} + v_z \frac{\partial v_z}{\partial z} \right) \\ = -\frac{\partial p}{\partial z} + \mu \left[ \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial v_z}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 v_z}{\partial \theta^2} + \frac{\partial^2 v_z}{\partial z^2} \right] + \rho f_z \\ \frac{1}{r} \frac{\partial (r v_r)}{\partial r} + \frac{1}{r} \frac{\partial v_\theta}{\partial \theta} + \frac{\partial v_z}{\partial z} = 0 \end{aligned}$$

The stress tensor (9.1) for a Newtonian fluid is

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{rr} & \sigma_{r\theta} & \sigma_{rz} \\ \sigma_{\theta r} & \sigma_{\theta\theta} & \sigma_{\theta z} \\ \sigma_{zr} & \sigma_{z\theta} & \sigma_{zz} \end{pmatrix},$$

where  $\sigma_{rr} = -p + 2\mu \partial_r v_r$ ,  $\sigma_{r\theta} = \sigma_{\theta r} = \mu \left( r \partial_r \left( \frac{1}{r} v_\theta \right) + \frac{1}{r} \partial_\theta v_r \right)$ ,  $\sigma_{rz} = \sigma_{rz} = \mu (\partial_z v_r + \partial_r v_z)$ ,  $\sigma_{\theta\theta} = -p + 2\mu \frac{1}{r} (\partial_\theta v_\theta + v_r)$ ,  $\sigma_{z\theta} = \sigma_{\theta z} = \mu (\partial_z v_\theta + \frac{1}{r} \partial_\theta v_z)$ , and  $\sigma_{zz} = -p + 2\mu \partial_z v_z$ .

Transformation laws for velocities:

$$\begin{aligned} u &= v_r \cos \theta - v_\theta \sin \theta & v_r &= \frac{1}{\sqrt{x^2 + y^2}} (xu + yv) \\ v &= v_r \sin \theta + v_\theta \cos \theta & v_\theta &= \frac{1}{\sqrt{x^2 + y^2}} (-yu + xv) \\ w &= v_z & v_z &= w. \end{aligned}$$

Transformation laws for derivatives:

$$\begin{aligned} \frac{\partial}{\partial x} &= \cos \theta \frac{\partial}{\partial r} - \frac{\sin \theta}{r} \frac{\partial}{\partial \theta} & \frac{\partial}{\partial r} &= \frac{x}{\sqrt{x^2 + y^2}} \frac{\partial}{\partial x} + \frac{y}{\sqrt{x^2 + y^2}} \frac{\partial}{\partial y} \\ \frac{\partial}{\partial y} &= \sin \theta \frac{\partial}{\partial r} + \frac{\cos \theta}{r} \frac{\partial}{\partial \theta} & \frac{\partial}{\partial \theta} &= -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} \end{aligned}$$

$$\frac{\partial}{\partial z} = \frac{\partial}{\partial z} \qquad \frac{\partial}{\partial z} = \frac{\partial}{\partial z}$$

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + v_r \frac{\partial}{\partial r} + \frac{v_\theta}{r} \frac{\partial}{\partial \theta} + v_z \frac{\partial}{\partial z}.$$

Formulas from vector analysis:

$$\nabla \times \mathbf{v} = \left( \frac{1}{r} \frac{\partial v_z}{\partial \theta} - \frac{\partial v_\theta}{\partial z} \right) \mathbf{e}_r + \left( \frac{\partial v_r}{\partial z} - \frac{\partial v_z}{\partial r} \right) \mathbf{e}_\theta + \frac{1}{r} \left( \frac{\partial(r v_\theta)}{\partial r} - \frac{\partial v_r}{\partial \theta} \right) \mathbf{e}_z$$

$$\nabla^2 \phi = \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial \phi}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 \phi}{\partial \theta^2} + \frac{\partial^2 \phi}{\partial z^2}$$

$$\nabla \phi = \frac{\partial \phi}{\partial r} \mathbf{e}_r + \frac{1}{r} \frac{\partial \phi}{\partial \theta} \mathbf{e}_\theta + \frac{\partial \phi}{\partial z} \mathbf{e}_z$$

$$\nabla \cdot \mathbf{v} = \frac{1}{r} \frac{\partial(r v_r)}{\partial r} + \frac{1}{r} \frac{\partial v_\theta}{\partial \theta} + \frac{\partial v_z}{\partial z}.$$

# References

- AAVSO, American Association of Variable Star Observers. Website (2018), <http://www.aavso.org>
- E. Abraham, O. Penrose, Physics of negative absolute temperatures. *Phys. Rev. E* **95**, 012125 (2017). <https://doi.org/10.1103/PhysRevE.95.012125>. <https://link.aps.org/doi/10.1103/PhysRevE.95.012125>
- D. Agnani, P. Acharya, E. Martinez, T.T. Tran, F. Abraham, F. Tobin, J. Bentz, Fitting the elementary rate constants of the P-gp transporter network in the hMDR1-MDCK confluent cell monolayer using a particle swarm algorithm. *PLoS ONE* **6**(10), e25086-1–e25086-11 (2011)
- T. Alazard, Low Mach number limit of the full Navier-Stokes equations. *Arch. Ration. Mech. Anal.* **180**(1), 1–73 (2006). <https://doi.org/10.1007/s00205-005-0393-2>. ISSN 1432-0673
- G.L. Aranovich, M.D. Donohue, Eliminating the mean-free-path inconsistency in classical phenomenological model of diffusion for fluids. *Physica A* **373**, 119–141 (2007)
- R. Aris, Review of rational thermodynamics. *Am. Math. Mon.* **94**, 562–564 (1987)
- R. Aris, *Vectors, Tensors and the Basic Equations of Fluid Mechanics* (Dover, New York, 1990)
- T. Asai, K. Seo, O. Kobayashi, R. Sakashita, Fundamental aerodynamics of the soccer ball. *Sports Eng.* **10**, 101–110 (2007)
- C. Atkinson, G.E.H. Reuter, C.J. Ridler-Rowe, Traveling wave solution for some nonlinear diffusion equations. *SIAM J. Math. Anal.* **12**(6), 880–892 (1981). <https://doi.org/10.1137/0512074>
- J.S. Bader, R.W. Hammond, S.A. Henck, M.W. Deem, G.A. McDermott, J.M. Bustillo, J.W. Simpson, G.T. Mulhern, J.M. Rothberg, DNA transport by a micromachined Brownian ratchet device. *Proc. Natl. Acad. Sci.* **96**, 13165–13169 (1999)
- R.W. Balluffi, S.M. Allen, W.C. Carter, *Kinetics of Materials* (Wiley, New York, 2005)
- J.R. Bamforth, S. Kalliadasis, J.H. Merkin, S.K. Scott, Modelling flow-distributed oscillations in the CDIMA reaction. *Phys Chem Chem Phys* **2**, 4013–4021 (2000)
- J. Bang, R. Pan, T.M. Hoang, J. Ahn, C. Jarzynski, H.T. Quan, T. Li, Experimental realization of Feynman’s ratchet. *N. J. Phys.* **20**(10), 103032 (2018), <http://stacks.iop.org/1367-2630/20/i=10/a=103032>
- R.C. Batra, Universal relations for transversely isotropic elastic materials. *Math. Mech. Solids* **7**, 421–437 (2002)
- A. Bayon, F. Gascon, A. Varade, Measurement of the longitudinal and transverse vibration frequencies of a rod by speckle interferometry. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **40**, 265–269 (1993)
- A. Bayon, A. Varade, F. Gascon, Determination of the elastic constants of isotropic solids by optical heterodyne interferometry. *J. Acoust. Soc. Am.* **96**, 2589–2592 (1994)

- H. Bhatia, G. Norgard, V. Pascucci, and P.T. Bremer, The Helmholtz-Hodge decomposition - a survey. *IEEE Trans. Vis. Comput. Graph.* **19**(8), 1386–1404 (2013). <https://doi.org/10.1109/TVCG.2012.316>. ISSN 1077-2626
- T.A. Blackledge, C.Y. Hayashi, Silken toolkits: biomechanics of silk fibers spun by the orb web spider *Argiope argentata* (Fabricius 1775). *J Exp Biology* **209**, 2452–2461 (2006)
- J. Bluck, NASA tunnels test tennis balls. Press Release, 00-58AR (2000), [http://www.nasa.gov/centers/ames/news/releases/2000/00\\_58AR.html](http://www.nasa.gov/centers/ames/news/releases/2000/00_58AR.html)
- J. Blum, S. Bruns, D. Rademacher, A. Voss, B. Willenberg, M. Krause, Measurement of the translational and rotational brownian motion of individual particles in a rarefied gas. *Phys. Rev. Lett.* **97**, 230601 (2006)
- G.W. Bluman, S. Anco, *Symmetry and Integration Methods for Differential Equations* (Springer, New York, 2002)
- G.W. Bluman, A. Cheviakov, S. Anco, *Applications of Symmetry Methods to Partial Differential Equations* (Springer, New York, 2010)
- R.A. Blythe, M.R. Evans, Nonequilibrium steady states of matrix-product form: a solver's guide. *J. Phys. A Math. Theor.* **40**(46), R333 (2007). <http://stacks.iop.org/1751-8121/40/i=46/a=R01>
- C. Booth, T. Beer, J.D. Penrose, Diffusion of salt in tap water. *Am. J. Phys.* **46**, 525–527 (1978)
- M. Braun, *Differential Equations and Their Applications: An Introduction to Applied Mathematics*, 4th edn. (Springer, New York, 1993)
- G.E. Briggs, J.B.S. Haldane, A note on the kinetics of enzyme action. *Biochem. J.* **19**, 338–339 (1928)
- B. Brixner, Trinity: 16 July 1945. Website (2018), [http://www.radiochemistry.org/history/nuke\\_tests/trinity/index.html](http://www.radiochemistry.org/history/nuke_tests/trinity/index.html)
- F.N.M. Brown, See the wind blow. Dept. Aerosp. Mech. Eng. Rep., Univ. of Notre Dame (1971)
- J.C. Butcher, *Numerical Methods for Ordinary Differential Equations*, 3rd edn. (Wiley, New York, 2016)
- J. Carlson, A. Jaffe, A. Wiles, *The Millennium Prize Problems* (American Mathematical Society, Providence, 2006). ISBN 9780821836798
- L. Cavaleri, J.-H.G.M. Alves, F. Arduin, A. Babanin, M. Banner, K. Belibassakis, M. Benoit, M. Donelan, J. Groeneweg, T.H.C. Herbers, P. Hwang, P.A.E.M. Janssen, T. Janssen, I.V. Lavrenov, R. Magne, J. Monbaliu, M. Onorato, V. Polnikov, D. Resio, W.E. Rogers, A. Sheremet, J. McKee Smith, H.L. Tolman, G. van Vledder, J. Wolf, I. Young, Wave modelling - the state of the art. *Prog. Oceanogr.* **75**(4), 603–674 (2007). <https://doi.org/10.1016/j.pcean.2007.05.005>. <http://www.sciencedirect.com/science/article/pii/S0079661107001206>. ISSN 0079-6611
- T. Cebeci, J. Cousteix, *Modeling and Computation of Boundary-layer Flows*, 2nd edn. (Springer, New York, 2005)
- L.-Q. Chen, H. Ding, Two nonlinear models of a transversely vibrating string. *Arch. Appl. Mech.* **78**(5), 321–328 (2008). <https://doi.org/10.1007/s00419-007-0164-7>. ISSN 1432-0681
- D. Chiron, Travelling waves for the nonlinear Schrödinger equation with general nonlinearity in dimension one. *Nonlinearity* **25**(3), 813 (2012). <http://stacks.iop.org/0951-7715/25/i=3/a=813>
- R.C.V. Coelho, M. Mendoza, M.M. Doria, H.J. Herrmann, Kelvin-Helmholtz instability of the Dirac fluid of charge carriers on graphene. *Phys Rev B* **96**, 184307 (2017). <https://doi.org/10.1103/PhysRevB.96.184307>. <https://link.aps.org/doi/10.1103/PhysRevB.96.184307>
- A. Colagrossi, D. Durante, J.B. Bonet, A. Souto-Iglesias, Discussion of Stokes' hypothesis through the smoothed particle hydrodynamics model. *Phys. Rev. E* **96**, 023101 (2017) <https://doi.org/10.1103/PhysRevE.96.023101>. <https://link.aps.org/doi/10.1103/PhysRevE.96.023101>
- M.S. Cramer, Numerical estimates for the bulk viscosity of ideal gases. *Phys Fluids* **24**(6), 066102 (2012). <https://doi.org/10.1063/1.4729611>
- M. Destrade, P.A. Martin, T.C.T. Ting, The incompressible limit in linear anisotropic elasticity, with applications to surface waves and elastostatics. *J. Mech. Phys. Solids* **50**(7), 1453–1468 (2002). [https://doi.org/10.1016/S0022-5096\(01\)00121-1](https://doi.org/10.1016/S0022-5096(01)00121-1). <http://www.sciencedirect.com/science/article/pii/S0022509601001211>. ISSN 0022-5096

- P.G. Drazin, W.H. Reid, *Hydrodynamic Stability*, 2nd edn. (Cambridge University Press, Cambridge, 2004)
- D. Drew, *Traffic Flow Theory and Control* (McGraw Hill, New York, 1968)
- A.S. Dukhin, P.J. Goetz, Bulk viscosity and compressibility measurement using acoustic spectroscopy. *J. Chem. Phys.* **130**(12), 124519 (2009). <https://doi.org/10.1063/1.3095471>
- B. Eckhardt, T.M. Schneider, B. Hof, J. Westerweel, Turbulence transition in pipe flow. *Annu. Rev. Fluid Mech.* **39**, 447–468 (2007)
- C.J. Efthimiou, M.D. Johnson, Domino waves. *SIAM Rev.* **49**, 111–120 (2007)
- G.A. El, M.A. Hoefer, M. Shearer, Dispersive and diffusive-dispersive shock waves for nonconvex conservation laws. *SIAM Rev.* **59**(1), 3–61 (2017). <https://doi.org/10.1137/15M1015650>
- O. el Moctar, University of Duisburg-Essen. Website (2018), <https://www.uni-due.de/ISMT/>
- J. Ellenberger, P.J. Klijn, M. Tels, J. Vleggaar, Construction and performance of a cone-and-plate rheogoniometer with air bearings. *J. Phys. E Sci. Instrum.* **9**, 763–765 (1976)
- R. Engbert, F. Drepper, Chance and chaos in population biology, models of recurrent epidemics and food chain dynamics. *Chaos Solutions Fractals* **4**, 1147–1169 (1994)
- A.C. Eringen, *Microcontinuum Field Theories II. Fluent Media* (Springer, New York, 2001)
- ESO/L. Calçada/M. Kornmesser, Massive star birth (2019), <https://luiscalcada.com/massive-star-birth>. Accessed 10 Jan 2019
- L.C. Evans, *Partial Differential Equations*. Graduate Studies in Mathematics, vol. 19, 2nd edn. (American Mathematical Society, Providence, 2010)
- G. Eyink, U. Frisch, R. Moreau, A. Sobolevski, Euler equations: 250 years on, proceedings of an international conference. *Physica D* **237**(14–17), 1825–2250 (2008)
- I. Famili, B.O. Palsson, The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. *Biophys. J.* **85**(1), 16–26 (2003)
- R.P. Feynman, R.B. Leighton, M. Sands, *The Feynman Lectures on Physics, Vol. 1*, 2nd edn. (Addison Wesley, Reading, 2005)
- A. Fick, On liquid diffusion. *Philos. Mag.* **10**, 31–39 (1885)
- R.J. Field, R.M. Noyes, Oscillations in chemical systems. IV. Limit cycle behavior in a model of a real chemical reaction. *J. Am. Chem. Soc.* **60**, 1877–1884 (1974)
- R.J. Field, E. Koros, R.M. Noyes, Oscillations in chemical systems. II. Thorough analysis of temporal oscillation in the bromate-cerium-malonic acid system. *J. Am. Chem. Soc.* **94**, 8649–8664 (1972)
- M. Finnis, *Interatomic Forces in Condensed Matter* (Oxford University Press, Oxford, 2010)
- M. Frewer, More clarity on the concept of material frame-indifference in classical continuum mechanics. *Acta Mech.* **202**, 213–246 (2009)
- A. Friedman, *Generalized Functions and Partial Differential Equations* (Dover, New York, 2005)
- Y.C. Fung, *Biomechanics: Mechanical Properties of Living Tissues*, 2nd edn. (Springer, New York, 1993)
- I. Gasser, On non-entropy solutions of scalar conservation laws for traffic flow. *ZAMM* **83**, 137–143 (2003)
- F. Giustino, *Materials Modelling Using Density Functional Theory* (Oxford University Press, Oxford, 2014)
- O. Golinelli, K. Mallick, The asymmetric simple exclusion process: an integrable model for non-equilibrium statistical mechanics. *J. Phys. A Math. Gen.* **39**(41), 12679 (2006). <http://stacks.iop.org/0305-4470/39/i=41/a=S03>
- S.R. Goodwill, S.B. Chin, and S.J. Haake, Aerodynamics of spinning and non-spinning tennis balls. *J. Wind Eng. Ind. Aerodyn.* **92**, 935–958 (2004)
- P. Gray, S.K. Scott, *Chemical Oscillations and Instabilities: Non-linear Chemical Kinetics*. (Oxford University Press, Oxford, 1994)
- H. Greenberg, An analysis of traffic flow. *Oper. Res.* **7**, 79–85 (1959)
- R.D. Gregory, Helmholtz's theorem when the domain is infinite and when the field has singular points. *Q. J. Mech. Appl. Math.* **49**, 439–450 (1996)



- D.F. Griffiths, D.J. Higham, *Numerical Methods for Ordinary Differential Equations: Initial Value Problems*. Springer Undergraduate Mathematics Series (Springer, New York, 2010). ISBN 9780857291486
- G.W. Griffiths, W.E. Schiesser, Linear and nonlinear waves. *Scholarpedia* **4**(7), 4308 (2009). <https://doi.org/10.4249/scholarpedia.4308>. revision #154041
- E.-Y. Guo, H.-X. Xie, S.S. Singh, A. Kirubanandham, T. Jing, N. Chawla, Mechanical characterization of microconstituents in a cast duplex stainless steel by micropillar compression. *Mater. Sci. Eng. A Struct. Mater.* **598**(Supplement C), 98–105 (2014). <https://doi.org/10.1016/j.msea.2014.01.002>. <http://www.sciencedirect.com/science/article/pii/S0921509314000100>. ISSN 0921-5093
- M.E. Gurtin, L.C. Martins, Cauchy's theorem in classical physics. *Arch. Ration. Mech. Anal.* **60** (4), 305–324 (1976). <https://doi.org/10.1007/BF00248882>. ISSN 1432-0673
- M.E. Gurtin, V.J. Mizel, W.O. Williams, A note on Cauchy's stress theorem. *J. Math. Anal. Appl.* **22**(2), 398–401 (1968). [https://doi.org/10.1016/0022-247X\(68\)90181-9](https://doi.org/10.1016/0022-247X(68)90181-9). <http://www.sciencedirect.com/science/article/pii/0022247X68901819>. ISSN 0022-247X
- J.K. Hale, H. Kocak, *Dynamics and Bifurcations* (Springer, New York, 1996)
- P. Hänggi, F. Marchesoni, Artificial Brownian motors: Controlling transport on the nanoscale. *Rev. Mod. Phys.* **81**, 387–442 (2009). <https://doi.org/10.1103/RevModPhys.81.387>. <https://link.aps.org/doi/10.1103/RevModPhys.81.387>
- V. Henri, *Lois générales de l'action des diastases* (Librairie Scientifique A. Hermann, Paris, 1903)
- N.E. Henriksen, F.Y. Hansen, *Theories of Molecular Reaction Dynamics: The Microscopic Foundation of Chemical Kinetics* (Oxford University Press, Oxford, 2008)
- P.V. Hobbs, A.J. Kezweent, Splashing of a water drop. *Exp. Fluids* **155**, 1112–1114 (1967)
- P.M. Hoffmann, How molecular motors extract order from chaos (a key issues review). *Rep. Prog. Phys.* **79**(3), 032601 (2016). <http://stacks.iop.org/0034-4885/79/i=3/a=032601>
- M.H. Holmes, Finite deformation of soft tissue: Analysis of a mixture model in uni-axial compression. *J. Biomech. Eng.* **108**(4), 372–381 (1986)
- M.H. Holmes, *Introduction to Numerical Methods in Differential Equations*. Texts in Applied Mathematics, vol. 52 (Springer, New York, 2005)
- M. Holmes, Asymmetric random walks and drift-diffusion. *Europhys. Lett.* **102**(3), 30005 (2013a). <http://stacks.iop.org/0295-5075/102/i=3/a=30005>
- M.H. Holmes, *Introduction to Perturbation Methods*. Texts in Applied Mathematics, vol. 20, 2nd edn. (Springer, New York, 2013b)
- M.H. Holmes, *Introduction to Scientific Computing and Data Analysis*. Texts in Computational Science and Engineering, vol. 13 (Springer, New York, 2016)
- M.H. Holmes, V.C. Mow, W.M. Lai, The nonlinear interaction of solid and fluid in the creep response of articular cartilage. *Biorheology* **20**, 422 (1983)
- F. Horn, R. Jackson, General mass action kinetics. *Arch. Ration. Mech. Anal.* **47**(2), 81–116 (1972). <https://doi.org/10.1007/BF00251225>. ISSN 1432-0673
- P.L. Houston, *Chemical Kinetics and Reaction Dynamics* (Dover, New York, 2006)
- K. Hutter, K. Johnk, *Continuum Methods of Physical Modeling* (Springer, New York, 2004)
- P.E. Hydon, *Symmetry Methods for Differential Equations: A Beginner's Guide*. Cambridge Texts in Applied Mathematics (Cambridge University Press, Cambridge, 2000). ISBN 9780521497862
- U. Irion, A.P. Singh, C. Nusslein-Volhard, The developmental genetics of vertebrate color pattern formation: Lessons from zebrafish, in *Essays on Developmental Biology, Part B*, ed. by P.M. Wassarman. Current Topics in Developmental Biology, vol. 117 (Academic, New York, 2016), pp. 141–169. <https://doi.org/10.1016/bs.ctdb.2015.12.012>. <http://www.sciencedirect.com/science/article/pii/S0070215315002197>
- P. Iskyan, Cloud formation over Breckenridge, Colorado (2019). [https://www.facebook.com/piskyan/media\\_set?set=a.1777219625455&type=3](https://www.facebook.com/piskyan/media_set?set=a.1777219625455&type=3). Accessed 10 Jan 2019
- D.D. Joseph, Potential flow of viscous fluids: Historical notes. *Int. J. Multiphase Flow* **32**, 285–310 (2006)
- M. Kac, Can one hear the shape of a drum? *Am. Math. Mon.* **73**, 1–23 (1966)

- N.G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd edn. (North-Holland, Amsterdam, 2007)
- S. Kavosi, *Kelvin-Helmholtz Instability at Earth's magnetopause: THEMIS Observations and OpenGGCM Simulations*, PhD thesis, University of New Hampshire (2015)
- J.B. Keller, Diffusion at finite speed and random walks. *Proc. Natl. Acad. Sci.* **101**, 1120–1122 (2004)
- D. Kiener, W. Grosinger, G. Dehm, R. Pippan, A further step towards an understanding of size-dependent crystal plasticity: in situ tension experiments of miniaturized single-crystal copper samples. *Acta Mater.* **56**(3), 580–592 (2008). <https://doi.org/10.1016/j.actamat.2007.10.015>. <http://www.sciencedirect.com/science/article/pii/S1359645407006969>. ISSN 1359-6454
- J.K. Knowles, On entropy conditions and traffic flow models. *ZAMM* **88**, 64–73 (2008)
- S.V. Kryatov, E.V. Rybak-Akimova, A.Y. Nazarenko, P.D. Robinson, A dinuclear iron(III) complex with a bridging urea anion: implications for the urease mechanism. *Chem. Commun.* **11**, 921–922 (2000)
- C. Kunkle, Velocity field in a pipe, in *Millersville University Physics: Experiment of the Month* (2008). <https://www.millersville.edu/physics/experiments/067/index.php>
- M. Kwan, *A Finite Deformation Theory for Nonlinearly Permeable Cartilage and Other Soft Hydrated Connective Tissues and Rheological Study of Cartilage Proteoglycans*, PhD Thesis, RPI (1985)
- R.S. Lakes, Viscoelastic measurement techniques. *Rev. Sci. Instrum.* **75**, 797–810 (2004)
- B. Lau, O. Kedem, J. Schwabacher, D. Kwasnieski, E.A. Weiss, An introduction to ratchets in chemistry and biology. *Mater. Horiz.* **4**, 310–318 (2017). <https://doi.org/10.1039/C7MH00062F>
- E. Lauga, M.P. Brenner, H.A. Stone, Microfluidics: the no-slip boundary condition, in *Handbook of Experimental Fluid Dynamics*, ed. by C. Tropea, A.L. Yarin, J.F. Foss (Springer, New York, 2007)
- P.D. Lax, *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves* (Society for Industrial and Applied Mathematics, Philadelphia, 1973). <https://doi.org/10.1137/1.9781611970562>. <https://epubs.siam.org/doi/abs/10.1137/1.9781611970562>
- H.G. Lee, J. Kim, Two-dimensional Kelvin–Helmholtz instabilities of multi-component fluids. *Eur. J. Mech. B Fluids* **49**, 77–88 (2015). <https://doi.org/10.1016/j.euromechflu.2014.08.001>. <http://www.sciencedirect.com/science/article/pii/S0997754614001319>. ISSN 0997-7546
- D.S. Lemons, A. Gythiel, Paul Langevin's 1908 paper, 'on the theory of Brownian motion'. *Am. J. Phys.* **65**, 1079–1081 (1997)
- Y.K. Leong, Y.L. Yeow, Obtaining the shear stress shear rate relationship and yield stress of liquid foods from couette viscometry data. *Rheol. Acta* **42**, 365–371 (2003)
- B.W. Levin, M.A. Nosov, *Physics of Tsunamis* (Springer, New York, 2016)
- K.M. Liew, B.J. Chen, Z.M. Xiao, Analysis of fracture nucleation in carbon nanotubes through atomistic-based continuum theory. *Phys Rev B: Condens. Matter* **71**, 235424 (2005)
- M.J. Lighthill, G.B. Whitham, On kinematic waves; II. A theory of traffic flow on long crowded roads. *Proc. R. Soc. Lond. Ser. A* **229A**, 317–345 (1955)
- K. Lucas, *Molecular Models for Fluids* (Cambridge University Press, Cambridge, 2007)
- Google Maps, Map of Arlington Memorial Bridge. Website (2007), <http://maps.google.com/>
- J.E. Mark, B. Erman, *Rubberlike Elasticity: A Molecular Primer*, 2nd edn. (Cambridge University Press, Cambridge, 2007)
- J.A. Maroto, J. Duenas-Molina, J. de Dios, Experimental evaluation of the drag coefficient for smooth spheres by free fall experiments in old mines. *Eur. J. Phys.* **26**, 323–330 (2005)
- J.E. Marsden, T.J.R. Hughes, *Mathematical Foundations of Elasticity* (Dover, New York, 1994)
- R.M. Mazo, *Brownian Motion: Fluctuations, Dynamics, and Applications* (Oxford University Press, Oxford, 2002)
- R.D. Mehta, Aerodynamics of sports balls. *Ann. Rev. Fluid Mech.* **17**, 151–189 (1985)

- H. Meinhardt, Turing's theory of morphogenesis of 1952 and the subsequent discovery of the crucial role of local self-enhancement and long-range inhibition. *Interface Focus* **2**(4), 407–416 (2012). <https://doi.org/10.1098/rsfs.2011.0097>. <http://rsfs.royalsocietypublishing.org/content/2/4/407>. ISSN 2042-8898
- L. Michaelis, M. Menten, Die kinetik der invertinwirkung. *Biochem Z* **49**, 333–369 (1913)
- S.G. Mikhlin, *Mathematical Physics, an Advanced Course* (North-Holland, Amsterdam, 1970)
- P. Moon, D.E. Spencer, *Field Theory Handbook: Including Coordinate Systems, Differential Equations and Their Solutions* (Springer, New York, 1988)
- Y. Morita, N. Tomita, H. Aoki, S. Wakitani, Y. Tamada, T. Suguro, K. Ikeuchi, Visco-elastic properties of cartilage tissue regenerated with fibroin sponge. *Bio-Med. Mater. Eng.* **12**, 291–298 (2002)
- S. Morris, Boeing 777-236/ER aircraft, in *AirTeamImages* (2006)
- J. Mueller, S. Siltanen, *Linear and Nonlinear Inverse Problems with Practical Applications* (Society for Industrial and Applied Mathematics, Philadelphia, 2012). <https://doi.org/10.1137/1.9781611972344>. <http://epubs.siam.org/doi/abs/10.1137/1.9781611972344>
- T. Mullin, Experimental studies of transition to turbulence in a pipe. *Annu. Rev. Fluid Mech.* **43** (1), 1–24 (2011). <https://doi.org/10.1146/annurev-fluid-122109-160652>
- A. Murdoch, Some primitive concepts in continuum mechanics regarded in terms of objective space-time molecular averaging: the key role played by inertial observers. *J. Elast.* **84**, 69–97 (2006)
- K. Nagel, M. Schreckenberg, A cellular automaton model for freeway traffic. *J. Phys. I France* **2** (12), 2221–2229 (1992). <https://doi.org/10.1051/jpl:1992277>
- Ames Research Center NASA, Wind tunnel images. Website (2018), <https://www.hq.nasa.gov/office/aero/aavp/aetc/test/process-improvement-gallery.html>
- G.F. Newell, Nonlinear effects in the dynamics of car following. *Oper. Res.* **9**, 209–229 (1961)
- J. Nordström, A roadmap to well posed and stable problems in computational physics. *J. Sci. Comput.* **71**(1), 365–385 (2017). <https://doi.org/10.1007/s10915-016-0303-9>
- B.K. Oksendal, *Stochastic Differential Equations: An Introduction with Applications*, 6th edn. (Springer, New York, 2003)
- A. Paquier, F. Moisy, M. Rabaud, Viscosity effects in wind wave generation. *Phys. Rev. Fluids* **1**, 083901 (2016). <https://doi.org/10.1103/PhysRevFluids.1.083901>. <https://link.aps.org/doi/10.1103/PhysRevFluids.1.083901>
- R. Penrose, *The Road to Reality: A Complete Guide to the Laws of the Universe* (Vintage Books, New York, 2007)
- M. Polettini, M. Esposito, Irreversible thermodynamics of open chemical networks. i. emergent cycles and broken conservation laws. *J. Chem. Phys.* **141**(2), 024117 (2014). <https://doi.org/10.1063/1.4886396>
- Y. Qi, T. Cagin, Y. Kimura, W.A. Goddard, Molecular-dynamics simulations of glass formation and crystallization in binary liquid metals: Cu-Ag and Cu-Ni. *Phys. Rev. B* **59**, 3527–3533 (1999). <https://doi.org/10.1103/PhysRevB.59.3527>. <https://link.aps.org/doi/10.1103/PhysRevB.59.3527>
- H. Rakha, M. Van Aerde, Calibrating steady-state traffic stream and car-following models using loop detector data. *Transp. Sci.* **44**(2), 151–168 (2010). <https://doi.org/10.1287/trsc.1090.0297>
- M. Ramírez-Escudero, M. Gimeno-Pérez, B. González, D. Linde, Z. Merdzo, M. Fernández-Lobato, J. Sanz-Aparicio, Structural analysis of  $\beta$ -fructofuranosidase from *Xanthophyllomyces dendrorhous* reveals unique features and the crucial role of N-glycosylation in oligomerization and activity. *J. Biol. Chem.* **291**(13), 6843–6857 (2016). <https://doi.org/10.1074/jbc.M115.708495>. <http://www.jbc.org/content/291/13/6843.abstract>
- P. Raos, Modelling of elastic behaviour of rubber and its application in FEA. *Plast. Rubber Compos. Process. Appl.* **19**, 293–303 (1993)
- C. Reder, Metabolic control theory: a structural approach. *J. Theor. Biol.* **135**(2), 175–201 (1988) [https://doi.org/10.1016/S0022-5193\(88\)80073-0](https://doi.org/10.1016/S0022-5193(88)80073-0). <http://www.sciencedirect.com/science/article/pii/S0022519388800730>. ISSN 0022-5193

- L. Reese, A. Melbinger, E. Frey, Crowding of molecular motors determines microtubule depolymerization. *Biophys. J.* **101**(9), 2190–2200 (2011). <https://doi.org/10.1016/j.bpj.2011.09.009>. <http://www.sciencedirect.com/science/article/pii/S0006349511010630>. ISSN 0006-3495
- P.I. Richards, Shock waves on the freeway. *Oper. Res.* **4**, 42–51 (1956)
- R. Rioboo, C. Bauthier, J. Conti, M. Voue, J. De Coninck, Experimental investigation of splash and crown formation during single drop impact on wetted surfaces. *Exp. Fluids* **35**, 648–652 (2003)
- R.S. Rivlin, J.L. Ericksen, Stress-deformation relations for isotropic materials. *J. Ration. Mech. Anal.* **4**, 323–425 (1955). <http://www.jstor.org/stable/24900365>. ISSN 19435282, 19435290
- W. Rudin, *Principles of Mathematical Analysis*, 3rd edn. (McGraw-Hill, New York, 1976)
- H. Sawada, T. Kunimasu, Sphere drag measurements with the NAL 60cm MSBS. *J. Wind Eng.* **98**, 129–136 (2004)
- S. Schochet, The incompressible limit in nonlinear elasticity. *Commun. Math. Phys.* **102**(2), 207–215 (1985). <https://doi.org/10.1007/BF01229377>. ISSN 1432-0916
- D. Schomburg, D. Stephan, *Enzyme Handbook* (Springer, New York, 1997)
- S. Schuster, T. Hofer, Determining all extreme semi-positive conservation relations in chemical reaction systems: a test criterion for conservativity. *J. Chem. Soc. Faraday Trans.* **87**, 2561–2566 (1991) <https://doi.org/10.1039/FT9918702561>
- L.A. Segel, M. Slemrod, The quasi-steady-state assumption: A case study in perturbation. *SIAM Rev.* **31**, 446–477 (1989)
- D. Shaw, Diffusion in Semiconductors, in *Springer Handbook of Electronic and Photonic Materials*, ed. by S. Kasap, P. Capper, Chapter 6 (Springer, New York, 2017), pp. 133–149
- Y.M. Shtemler, M. Mond, V. Cherniavskii, E. Golbraikh, Y. Nissim, An asymptotic model for the Kelvin–Helmholtz and Miles mechanisms of water wave generation by wind. *Phys. Fluids* **20**(9), 094106 (2008). <https://doi.org/10.1063/1.2980350>
- M.J. Skaug, C. Schwemmer, S. Fringes, C.D. Rawlings, A.W. Knoll, Nanofluidic rocking Brownian motors. *Science* **359**(6383), 1505–1508 (2018). <https://doi.org/10.1126/science.aal3271>. <http://science.sciencemag.org/content/359/6383/1505>. ISSN 0036-8075
- A.J. Smits, S. Ogg, *Aerodynamics of the Golf Ball*. Biomedical Engineering Principles in Sports (Kluwer Academic, Boston, 2004)
- C.G. Speziale, Comments on the material frame-indifference controversy. *Phys. Rev. A At. Mol. Opt. Phys.* **36**, 4522–4525 (1987)
- C.G. Speziale, A review of material frame-indifference in mechanics. *Appl. Mech. Rev.* **51**, 489–504 (1998)
- J.P. Steinbrenner, J.P. Abelanet, Anisotropic tetrahedral meshing based on surface deformation techniques, in *Proceedings of the AIAA 45th Aerospace Sciences Meeting*, Reno, NV (2007), p. AIAA–2007–0554
- S.H. Strogatz, *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering*, 2nd edn. (Westview Press, Cambridge, 2014)
- W.J. Stronge, D. Shu, The domino effect: Successive destabilization by cooperative neighbours. *Proc. R. Soc. A* **418**, 155–163 (1988)
- B. Svendsen, A. Bertram, On frame-indifference and form-invariance in constitutive theory. *Acta Mech.* **132**, 195–207 (1999)
- R.H. Swendsen, How physicists disagree on the meaning of entropy. *Am. J. Phys.* **79**(4), 342–348 (2011). <https://doi.org/10.1119/1.3536633>
- R.H. Swendsen, Thermodynamics of finite systems: a key issues review. *Rep. Prog. Phys.* **81**(7), 072001 (2018). <http://stacks.iop.org/0034-4885/81/i=7/a=072001>
- H.A. Tahini, A. Chroneos, S.C. Middleburgh, U. Schwingenschlogl, R.W. Grimes, Ultrafast palladium diffusion in germanium. *J. Mater. Chem. A* **3**, 3832–3838 (2015). <https://doi.org/10.1039/C4TA06210H>
- S. Taneda, Oscillation of the wake behind a flat plate parallel to the flow. *J. Phys. Soc. Jpn.* **13**(4), 418–425 (1958). <https://doi.org/10.1143/JPSJ.13.418>

- R. Temam, *Navier-Stokes Equations: Theory and Numerical Analysis* (American Mathematical Society, Providence, 2001)
- Thoisoi2, Chemical Clock, Briggs-Rauscher oscillating Reaction! Website (2014), <https://www.youtube.com/watch?v=WpBwlSn1XPQ>
- S.A. Thorpe, A method of producing a shear flow in a stratified fluid. *J. Fluid Mech.* **32**(4), 693–704 (1968). <https://doi.org/10.1017/S0022112068000972>
- N. Tillmark, P.H. Alfredsson, Experiments on transition in plane Couette flow. *J. Fluid Mech.* **235**, 89–102 (1992)
- T.T. Tran, A. Mittal, T. Aldinger, J.W. Polli, A. Ayrton, H. Ellens, J. Bentz, The elementary mass action rate constants of P-gp transport for a confluent monolayer of MDCKII-hMDR1 cells. *Biophys. J.* **88**, 715–738 (2005)
- C. Truesdell, *Rational Thermodynamics*, 2nd edn. (Springer, New York, 1984)
- A. Turing, The chemical basis of morphogenesis. *Philos. Trans. R. Soc. B* **237**, 37–72 (1952)
- O. Vallée, M. Soares, *Airy Functions and Applications to Physics*, 2nd edn. (Imperial College Press, London, 2010). <https://doi.org/10.1142/p345>. <https://www.worldscientific.com/doi/abs/10.1142/p345>
- H. van Haren, L. Gostiaux, E. Morozov, R. Tarakanov, Extremely long Kelvin-Helmholtz billow trains in the Romanche Fracture Zone. *Geophys. Res. Lett.* **41**(23), 8445–8451 (2014). <https://doi.org/10.1002/2014GL062421>. ISSN 1944-8007
- T. Vanderbilt, *Traffic: Why We Drive the Way We Do (and What It Says About Us)* (Knopf, New York, 2008)
- C.C. Wang, A new representation theorem for isotropic functions. *Arch. Ration. Mech. Anal.* **36** (3), 198–223 (1970). <https://doi.org/10.1007/BF00272242>. ISSN 1432-0673
- M. Watanabe, S. Kondo, Is pigment patterning in fish skin determined by the Turing mechanism? *Trends Genet.* **31**(2), 88–96 (2015). <https://doi.org/10.1016/j.tig.2014.11.005>. <http://www.sciencedirect.com/science/article/pii/S0168952514001978>. ISSN 0168-9525
- H.F. Weinberger, *A First Course in Partial Differential Equations: with Complex Variables and Transform Methods* (Dover, New York, 1995)
- A.P. Willis, J. Peixinho, R.R. Kerswell, T. Mullin, Experimental and theoretical progress in pipe flow transition. *Philos. Trans. A Math. Phys. Eng. Sci.* **366**(1876), 2671–2684 (2008). <https://doi.org/10.1098/rsta.2008.0063>. <http://rsta.royalsocietypublishing.org/content/366/1876/2671>. ISSN 1364-503X
- S.-H. Wu, N. Huang, E. Jaquay, M.L. Povinelli, Near-field, on-chip optical Brownian ratchets. *Nano Lett.* **16**(8), 5261–5266 (2016) <https://doi.org/10.1021/acs.nanolett.6b02426>. PMID: 27403605
- H. Xiao, O.T. Bruhns, A. Meyers, On isotropic extension of anisotropic constitutive functions via structural tensors. *ZAMM* **86**(2), 151–161 (2006). <https://doi.org/10.1002/zamm.200410226>. <https://onlinelibrary.wiley.com/doi/abs/10.1002/zamm.200410226>
- N. Zahibo, E. Pelinovsky, T. Talipova, A. Kozelkov, A. Kurkin, Analytical and numerical study of nonlinear effects at tsunami modeling. *Appl. Math. Comput.* **174**(2), 795–809 (2006). <https://doi.org/10.1016/j.amc.2005.05.014>. <http://www.sciencedirect.com/science/article/pii/S0096300305005102>. ISSN 0096-3003

# Index

## Symbols

$\ll$ , 56

$\llbracket \rrbracket$ , 4

## A

Admissibility condition, 267

Advection equation, 246

Airy function, 46

Alfven speed, 39

Almansi strain, 314, 342

Arrhenius equation, 113

Articular cartilage, 198, 313

Asymmetric simple exclusion process, 293

Asymptotically stable, 127, 129

Asymptotic expansion, 53, 65

Autocatalytic reaction, 112, 144

Avogadro's number, 177

## B

Balance law, 196, 238, 400

Balancing, 68, 72, 81

Bell curve, 169

Belousov-Zhabotinskii reaction, 143

Bernoulli's theorem, 460

Bessel function, 352, 376

Binding energy, 340

Blasius boundary layer, 474

Bobyreff-Forsyth formula, 490

Bohr radius, 39

Boltzmann constant, 177, 217

Boltzmann distribution, 203

Boundary layer coordinate, 71, 81, 473

Boundary layer solution, 71, 81

Boundary layer thickness, 470, 476

Bratu's equation, 101

Brittle material, 314

Brownian motion, 165

Brownian ratchet, 180

Buckingham Pi theorem, 16

Bungie cord, 99, 296, 317, 319, 333, 339

Burgers' equation, 40, 290

## C

Capture silk, 313

Carbon nanotube, 317, 341

Carburization, 178

Cauchy-Green deformation tensor, 336

Cauchy stress tensor, 406

Cellular automata modeling, 276

Characteristics, 249, 255

Clausius-Duhem inequality, 329

Complementary error function, 26, 183, 194, 352

Composite approximation, 74

Composite expansion, 83, 140

Compressive strain, 324, 347

Conservation law, 110, 120, 239

Constitutive law, 197, 312, 325

diffusion, 197

Greenshields law, 241, 257

linear elastic, 316, 345, 429

viscoelastic, 365

viscous fluid, 418

Contact discontinuity, 261

Continuity equation, 239  
 incompressible, 402  
 material, 306, 424  
 spatial, 306, 402

Control volume, 237, 337

Convolution theorem, 189, 356

Cooperativity, 162

Couette flow, 446

Creep, 312

## D

D'Alembert's paradox, 469

D'Alembert solution, 349

Deformation gradient, 336, 392, 427, 464

Density, 234, 305, 402

Descartes' rule of signs, 130, 163

Determinant, 413, 509

derivative, 398

Diffusion coefficient, 23, 175

Diffusion equation, 40, 175, 207  
 point source solution, 41, 206, 209, 455  
 radially symmetric, 210, 455

Diffusive boundary layer, 470

Dimensionally complete, 18, 20

Dimensionally homogeneous, 4, 17, 20

Dimensionless product, 8, 19  
 independent, 19

Dimension matrix, 17

Direct notation, 401

Displacement  
 gradient, 427  
 material, 296, 390  
 spatial, 297, 390

Distinguished limit, 175

Divergence Theorem, 399

Drag coefficient, 10

Drag on sphere, 6

Drift coefficient, 218

Drift diffusion, 202

Drift-diffusion equation, 224

Drift velocity, 225

Du Bois-Reymond lemma, 304

Ductile material, 314

Duffing equation, 99

## E

Einstein-Smoluchowski equation, 176, 221

Elastic beam, 43

Elastic limit, 324

Elastic modulus, 4, 316

Elastic string, 41

Elastomer, 324

Elementary reaction, 114, 149

Eley-Rideal mechanism, 152

Entropy, 4, 329

Epidemic equilibrium, 117, 158, 159

Error function, 352

Euclidean transformation, 327, 409

Euler equations, 461

Eulerian coordinates, 390

Eulerian strain, 314

Expansion fan, 40, 266

Exponential order, 354

Extension ratio, 313, 339

Extreme parameter value, 150

## F

Fick's law of diffusion, 197

First Piola-Kirchhoff stress tensor, 424

Fisher's equation, 32

Fixed junction model, 339

FKN mechanism, 143

Flux, 47, 195, 196, 236, 400

Form invariance, 410

Fourier law of heat conduction, 197

Fourier series, 503

Fourier transform, 186

Fracture, 324

Frame-indifference, 327, 410, 426

Free-surface problem, 478

Froude number, 492

Fundamental diagram, 244

Fundamental dimension, 3, 16

## G

Galilean transformation, 327, 409

Gap, 276

Geometric analysis, 124

Geometric Brownian motion, 219

Geometric linearity, 345, 363

Goldilocks, 175

Greenshields constitutive law, 241, 252

Green strain, 314, 428

## H

Half-plane of convergence, 354

Hanes-Woolf plot, 164

Heaviside step function, 353

Helical flow, 491

Helicity, 493

Helmholtz free energy, 329

Helmholtz Representation Theorem, 457, 474

Helmholtz's Third Vorticity Theorem, 465



Hencky strain, 314  
 Hill's equation, 162  
 Homogeneous material, 417  
 Hopf bifurcation, 132  
 Hurricane, 455  
 Hurwitz matrix, 163  
 Hydrogen-bromine reaction, 163  
 Hyperelasticity, 331, 332

## I

Ideal fluid, 461  
 Ideal gas, 332, 419  
 Impenetrability of matter, 303  
 Impermeability boundary condition, 421  
 Impulsive plate, 470, 493  
 Incompressibility  
   material coordinates, 439  
   spatial coordinates, 402  
 Indicator function, 190  
 Infinitesimal deformation, 363  
 Initial layer, 135  
 Inner solution, 71, 139  
 Instantaneous elastic modulus, 386  
 Integro-differential equation, 371, 375  
 Internal energy, 329  
 Interstitial diffusion, 177  
 Inverse Function Theorem, 304  
 Inverse problems, 362  
 Invertase, 116  
 Inviscid fluid, 461  
 Irrotational flow, 441, 457  
 Isotropic material, 413

## J

Jacobian matrix, 128, 392, 425  
 Jacobi's formula, 398  
 Jam density, 283

## K

Karman vortex street, 476  
 KdV equation, 46  
 Kelvin-Helmholtz instability, 484  
 Kelvin's Circulation Theorem, 464, 492  
 Kelvin's Minimum Energy Theorem, 441  
 Kelvin-Voigt model, 365  
 Kermack-McKendrick model, 104  
 Ketchup, 445, 449  
 Kinematic viscosity, 470  
 Kinetic energy, 328, 430, 433, 441  
 Kutta-Joukowski theorem, 469

## L

Lagrangian coordinates, 296, 390  
 Lagrangian strain, 314, 430  
 Lamè constants, 429  
 Langevin equation, 212  
 Laplace transform, 350  
 Laplacian, 418  
 Law of Mass Action, 108  
 Left Cauchy-Green deformation tensor, 433  
 Leibniz's rule, 304  
 Lengyel-Epstein model, 20  
 Lennard-Jones potential, 324  
 Limit cycle, 133  
 Lincoln Tunnel, 240  
 Lindeman model, 159  
 Linear flow, 435, 487  
 Linear stability analysis, 128  
 Logistic equation, 154  
 Luminosity, 35

## M

Magnetosonic waves, 39  
 Markovian forcing, 216  
 Markov property, 167  
 Mass density, 4  
 Mass, spring, dashpot, 41, 363  
 Master equation, 174, 208, 224  
 Matching condition, 73, 139  
 Material coordinate system, 296, 390  
 Material curve, 464  
 Material derivative, 301, 395  
 Material linearity, 345, 363, 385  
 Material velocity gradient tensor, 399  
 Maxwell model, 365  
 Mean free path, 171, 176  
 Mean-square displacement, 214, 220  
 Mechanical energy equation, 342, 430  
 Merge density, 235, 283  
 Merritt Parkway, 240  
 Metallic bonding, 321  
 Method of characteristics, 348  
   linear wave equation, 249  
   nonlinear wave equation, 255  
 Method of multiple scales, 89  
 Michaelis-Menten reaction, 115, 130  
 Midpoint strain, 314  
 Mobility, 202  
 Momentum equation  
   angular, 407, 424  
   material coordinates, 309, 337, 424  
   spatial coordinates, 309, 407  
 Mooney-Rivlin model, 341  
 Morse potential function, 341



**N**

Nagel-Schreckenberg model, 277  
 Navier equations, 430  
 Navier-Stokes equation, 418  
 Nernst-Planck law, 203  
 Neubert-Fung relaxation function, 373  
 Newtonian fluid, 416, 418, 445  
 Newton-Sefan law of cooling, 42, 94  
 Nondimensionalization, 27, 137  
 Non-isotropic material, 413  
 Non-Newtonian fluid, 448, 495  
 No-slip condition, 421, 462  
 Nuclear explosion, 38  
 Nullcline, 124, 146

**O**

Objective tensor, 410, 426  
 Oldroyd fluid, 386  
 One-way wave equation, 248  
 Oregonator, 144  
 Outer solution, 71, 81, 138, 145, 472  
 Overlap domain, 73

**P**

Partial derivative notation, 299, 500  
 Pascal, 317  
 Pathline, 468  
 Pauli exclusion principle, 322  
 Peanut butter, 445  
 Pendulum, 35, 84  
 P-glycoprotein, 134, 140  
 Phantom traffic jam, 254  
 Piecewise continuous, 503  
 Pipe flow, 35, 422, 450  
 Planck's constant, 39  
 Plane Couette flow, 447, 454  
 Plasticity, 324  
 Plug-flow reactor, 20  
 Point source solution of diffusion equation, 179, 206, 209, 455  
 Poiseuille flow, 422, 450, 486  
 Polar decomposition theorem, 428  
 Polyconvexity, 334  
 Polytropic fluid, 492  
 Potential energy, 328, 433  
 Potential flow, 466  
 Power-law fluid, 449, 495  
 Predator-prey model, 104, 118  
 Pressure, 332  
 Principal invariants, 413, 438  
 Principle of Dissipation, 330, 419

Principle of Material Frame-Indifference, 327, 409, 426  
 Projectile problem, 1, 27, 60  
 Pure shear, 434

**Q**

Quantum chromodynamics, 39  
 Quasi-steady state assumption, 135, 142

**R**

Radioactive decay, 103  
 Random walk, 167, 205  
     biased, 224  
     lazy, 222  
     non-rectangular lattice, 230  
     persistent, 222  
     with loss, 223  
     with memory, 222  
 Rankine-Hugoniot condition, 261, 386  
 Rarefaction wave, 266  
 Rate of deformation tensor, 416, 489  
 Reaction analysis, 123  
 Reaction-diffusion equations, 204  
 Red light - green light problem, 249, 279  
     modified, 256  
 Reduced entropy inequality, 330  
 Reduced problem, 30, 49  
 Reference configuration, 297, 391  
 Regular perturbation problem, 49  
 Reiner-Rivlin fluid, 416  
 Resonance, 361  
 Reynolds number, 10, 472  
 Reynolds Transport Theorem, 304, 397, 400  
 Riemann problem, 40, 263, 264  
 Right Cauchy-Green deformation tensor, 428  
 Rivlin-Ericksen representation theorem, 413, 417, 442  
 Rotation matrix, 394, 409, 428  
 Rozenzweig-MacArthur model, 161  
 Rubber, 313, 339

**S**

Scale functions, 65  
 Scale model testing, 12  
 Schnakenberg chemical oscillator, 160  
 Second law of thermodynamics, 329  
 Second Piola-Kirchhoff stress tensor, 427  
 Secular term, 88  
 Shear stress, 448, 449, 495  
 Shock wave, 261

Similarity variable, 24, 199, 210, 293  
 Simple extension, 440  
 Simple shear, 393  
 Singular perturbation problem, 66, 138  
 SIR model, 105  
     with vaccination, 158  
     with vital dynamics, 158  
 Slinky, 346, 357  
 Slip plane, 324  
 Small disturbance approximation, 253, 482  
 Space elevator, 339  
 Spatial coordinate system, 297, 390  
 Spatial velocity gradient tensor, 399  
 Spin tensor, 416  
 Standard linear model, 365  
 Steady flow, 446  
 Steady state, 111, 123, 311  
 Stirling's approximation, 172, 224  
 Stochastic differential equation, 212  
 Stoichiometric coefficients, 107, 119  
 Stoichiometric matrix, 120  
 Stokes drag formula, 12, 202, 216  
 Stokes-Einstein equation, 176, 217  
 Stokes' first problem, 470  
 Stokes flow, 11  
 Stokes hypothesis, 419  
 Stokes' Law, 37  
 Stored energy function, 433  
 Strain  
     Almansi, 314, 440, 441  
     energy function, 433  
     engineering, 314  
     Eulerian, 314  
     Green, 314, 338, 428  
     Hencky, 314, 339, 440, 441  
     Lagrangian, 314, 315, 430  
     midpoint, 314, 339  
     nominal, 314  
     tensor, 440  
     true, 314  
 Stream function, 474  
 Stress, 4, 307, 316, 403, 424  
 Stress power, 431  
 Stress relaxation, 311  
 Surface tension, 4, 494  
 Sutton-Chen potential, 324

## T

Tacoma Narrows Bridge, 362  
 Tautochrone problem, 384  
 Taylor-Couette problem, 489  
 Taylor-Sedov formula, 38  
 Taylor's theorem, 50, 497

Telegraph equation, 223  
 Temperature, 329, 419  
 Tensile strain, 324, 347  
 Tensor, 401  
 Toothpaste, 445, 449  
 Trace, 398, 413, 509  
 Traffic flow equation  
     linear, 241, 245  
     nonlinear, 242, 252, 270  
     small disturbance approximation, 253  
     wave velocity, 243, 252  
 Transcendentally small, 66  
 Trimerization, 162  
 Tsunami, 481  
 Two-timing, 89

## U

Uniform approximation, 74, 83  
 Uniform dilatation, 393  
 Universal gas constant, 177

## V

Van der Pol equation, 132  
 Van der Waals bonding, 324  
 Velocity  
     gradient tensor, 399, 415  
     material, 297, 390  
     spatial, 297, 390  
 Viscoelastic  
     fluid, 373  
     solid, 373  
 Viscoelasticity  
     Burger model, 381  
     creep function, 382  
     Kelvin-Voigt model, 365  
     Maxwell model, 365  
     relaxation function, 372  
     standard linear model, 365  
 Viscosity, 4, 332, 418, 446  
 Viscous dissipation function, 431, 489  
 Viscous fluid, 331  
 Volatility, 218  
 Volume fraction, 284  
 Vortex, 455  
     Burger's, 488  
     line, 457, 460  
     Oseen-Lamb, 456, 463  
     Taylor, 455, 487  
 Vorticity, 454, 462, 469, 489  
 Vorticity tensor, 416, 454  
 Vorticity transport equation, 488

**W**

Wave velocity, [243](#), [252](#)

Weak nonlinearity, [32](#)

Weber number, [36](#)

Webster's equation, [337](#)

Well-ordered, [65](#)

Well-ordering condition, [55](#)

**Y**

Young's modulus, [316](#), [345](#)